Use of Personality Assessment for the Prediction of Behaviour in Horses

Thesis

Use of Personality Assessment for the Prediction of Behaviour in Horses

Adele Sian Lloyd, BSc. (Hons), MSc.

A thesis in partial fulfilment of the requirements of the Open University for the degree of Doctor of Philosophy

Original submission: February 2007

Resubmitted: June 2008

The Open University

Harper Adams University College

Moulton College

Abstract

There are several potential applications for horse personality assessment, but first a reliable and valid form of assessment must be developed. The primary aim of this thesis was to develop a horse personality assessment method and test it for both reliability and validity by relating personality assessments to real-world observations. A 30-item rating questionnaire was developed and was named the Horse Personality Questionnaire (HPQ). This was used to assess 61 horses, each by three raters. 71.2% of horses and 25 of the items were rated consistently between raters. Principal component analysis (PCA) on these data extracted six components that were thought to describe horse personality; *Antagonism, Anxiousness, Activity, Protection, Sociability* and *Inquisitiveness*. Personality component scores were found to correlate with horse behaviours recorded in the field, thus demonstrating the reliability and validity of the HPQ. The HPQ was also used to explore breed differences in horse personality, with 1223 horses from eight different breeds assessed. The results provided strong evidence that horse breeds differ in personality, but those breeds with linked pedigrees or functions were shown to be more alike. These results suggested that personality in horses could be, to some extent, heritable and that humans have selected for different personality types. During the third study predictions of personality behaviour correlations were tested by first assessing 14 horses and then exposing each horse to three behaviour tests (learning, arena and turn-out tests). Of 25 predicted behaviours none were found to be significant. These results did not support previous indications that personality scores could be used to predict behaviour. It was concluded that the behaviour tests used and the predictions made may not have been suitable. The three experiments are discussed in terms of their implications for personality research and the potential applications of the HPQ in the equine industry. It was concluded that the HPQ was a reliable assessment method but required further development and testing prior to application in the equine industry.

Contents

List of Tables and Figures

Acknowledgements

There are many people that I am indebted to for their help and support throughout my Ph.D. Without them, I know I would not have been able to achieve what I have.

Firstly I am grateful for the help and support provided to me by my two key supervisors, Dr Jo Meredith and Dr Hannah Gauci. Jo helped build my knowledge of personality theory and also had great patience during the early stages of my Ph.D. when I was struggling to cope due to personal difficulties. When Jo went on maternity leave in 2005, Hannah took the reins and was the driving force I needed to get the thesis completed in good time. Both Hannah and Jo have been incredibly generous with their time with the provision of regular meetings and discussions as well as rapid and detailed feedback on the numerous drafts they have read. Their feedback and criticism has resulted in a vast improvement of my writing and research skills as well as increasing my confidence as a scientist. I would also like to thank Dr Robert Wilkinson, my Director of Studies at Harper Adams University College. Throughout the duration of my Ph.D. Robert has provided valuable advice and visited regularly. His experience of working with Ph.D. students and publishing journal articles has been most valuable.

I would also like to thank Alvaro Faria and colleagues at the Open University Statistics Department for their statistical support and advice. Alvaro has given a lot of his time in order to help me and his enthusiasm for my project was always very encouraging. I would also like to thank the two anonymous reviewers from Applied Animal Behaviour Science, whose generous comments on my first paper increased my confidence and enthusiasm dramatically. I am also grateful for the IT support provided by the Open University, Matthew Holden of Moulton College and Stuart Rimmell, formerly of Harper Adams.

I would like to thank the following people for their help during my first experiment; Fiona Gifkin at the Milton Keynes Equestrian Centre; Emma at East Lodge Farm, Sywell and Claire McClellan and Laura Oakley from Moulton College Equestrian Yard. I am very appreciative of the time that staff and owners at each of these establishments devoted to this study. Furthermore I am especially grateful to all the horse owners at Mallards Barn, Hannington, with particular thanks to Yvonne Smedley who helped to organise the involvement of this yard.

With regards to my second study, it would not have received such a large number of responses had it not been for the kind cooperation and support given by the breed societies and members involved in the survey. These were the Welsh Pony and Cob Society, Shetland Pony Studbook Society, Irish Draught Horse Society, Arab Horse Society, British Appaloosa Society and the Highland Pony Society. I would also like to thank Mary Davies and Christina Garside for their help in producing and printing the questionnaire. I am also grateful to Laura Hall for her help with publicity. Many thanks also to the owners, handlers and breeders of thoroughbred horses who spent time completing questionnaires often for multiple horses. Finally I am thankful for the encouraging emails and messages sent by respondents.

The third experiment was carried out at Moulton College Equestrian Yard and I am grateful for the help and time provided by the all the staff and students working on the yard during this experiment. I am also grateful to Rhiannon Kinsman and Lindsay Whistance for helping with these experiments, despite their own hectic schedules.

This project would not have been possible if it were not for the financial support provided to me. The Moulton College and the Thomas Harrison Trust provided partial funding as

well as office facilities and supervision. In addition to this I have received substantial financial support from my parents, without which I know I would have struggled to complete the thesis. The Association for the Study of Animal Behaviour (ASAB) have also contributed generously by providing financial support for me to attend three conferences.

Special thanks are also given to my Line Manager, Christine Oates, at Bishop Burton College. Christine allowed me to utilise a proportion of my working time to finish the final amendments of my thesis. Additionally, my colleagues (in particular, Angelique Mawdsley, Rebecca Smurthwaite and Melanie Watson-Smith) at Bishop Burton College have also provided support in a variety of different ways. This level of support during the last year was a great help in allowing me to complete my amendments within time.

Finally I would like to thank Lindsay Whistance, Cat Barlow, Peter Bond, my family and especially my mum (Anita Lloyd) for being good friends and always being there when things got tough. Although this thesis is my own work, it would not have been achievable without the help and support from all of these people, and for that I am very grateful.

1 Introduction

Recently, within animal science the study of personality has become an increasing area of interest (Gosling and John, 1999; Gosling, 2001; Gosling and Vazire, 2002). Research has begun to acknowledge that individuals in a population vary in their reactions to stimuli and, more importantly, that these reactions are consistent across time and situations (Gosling and John, 1999; Gosling, 2001; Gosling and Vazire, 2002). Non-human animal (henceforth referred to as animal) personality research has explored the development of reliable and valid assessment methods (e.g. Stevenson-Hinde *et al.,* 1980; Anderson, 1999; Momozawa *et al*., 2003), the biological controls of personality (e.g. Byrne and Suomi, 2002; Drent *et al*., 2002; Capitanio *et al*., 2004) and the application of personality assessment within the animal industry (e.g. Goddard and Beilharz, 1983, 1984; Visser *et al*., 2003a; Maejima *et al.,* in press). But what evidence is there to support claims of the existence of animal personality?

A wide range of evidence for animal personality was explored by Gosling and Vazire (2002) who described three criteria, adapted from Kenrick and Funder (1988), that should be met in order to establish the existence of personality traits and, therefore, demonstrate reliability of personality assessment methods. Their criteria focussed on: 1) agreement between raters; 2) use of personality ratings to predict behaviour; and 3) that ratings were shown to be true representations of an individual's personality. Through an extensive comparative evaluation of the available animal personality literature Gosling and Vazire (2002) demonstrated that animal personality research had produced results comparable to those found in human personality research. Furthermore, they identified that personality research that used rating data produced similar results to those that had used specific behavioural tests and observations. Gosling and Vazire (2002) concluded that, as a whole,

the research had met all of their criteria, but that individual studies had not necessarily met all three. In addition, they recommended that animal personality researchers should utilise these criteria in order validate their results and further strengthen the evidence for animal personality and our ability to assess it.

Animal personality has been explored in a wide range of species (for a review see Gosling and John, 1999) and this has, to some extent, included horses. Previous horse personality studies have, however generally focused on using behavioural tests (Le Scolan *et al.*, 1997; Wolff *et al.*, 1997; Anderson *et al.*, 1999; Visser *et al.*, 2001, 2003a; Seaman *et al.*, 2002; McCall *et al.*, 2006), but these can be restrictive in the elements of personality that they can measure. In contrast, in other animal personality studies researchers have developed rating questionnaires that can be completed by handlers and provide a detailed and broad representation of an individual's personality (for example see Stevenson-Hinde & Zunz, 1978; Stevenson-Hinde *et al.*, 1980; Gold & Maple, 1994; McGuire *et al.*, 1994; Dutton *et al.*, 1997; King & Figueredo, 1997; Morris *et al.*, 2002a, 2002b; Momozawa *et al.*, 2003, 2005). The resulting data are then often simplified into personality dimensions by using multivariate analysis techniques, such as principal components analysis (PCA). The simplifying of these data into components or personality dimensions allows for personality taxonomy to be explored. Comparative analyses can then be carried out, for example, Gosling and John (1999) compared the personality structures of a wide range of species and identified that there are key personality components that are recurrent across a range of species.

The Stevenson-Hinde *et al.* (1980) rhesus macaque (*Macaca mulatta)* rating questionnaire provides an example of a trait rating method that has been applied to a wide range of species. These have included cheetah (*Acinonyx jubatus*) (Wielebnowski, 1999), spotted

hyena (*Crocuta crocuta*) (Gosling, 1998) and chimpanzees (*Pan troglodytes*) (Martin, 2005), but this method has not yet been applied to horses. Instead, researchers that have used rating methods to assess horse personality have either developed novel assessment criteria (Creighton, 2003; Momozawa *et al.*, 2003; 2005, McGrogan, in press) or have adapted human personality questionnaires (Morris *et al.*, 2002a, 2002b). Although these studies have demonstrated individual differences in horse personality, there is little evidence that these studies have met all three of Gosling and Vazire's (2002) criteria. Future assessment methods should, therefore, be developed and evaluated with reference to these criteria. Furthermore, they should also allow for cross-species comparisons. The Stevenson-Hinde *et al.* (1980) trait list, therefore, demonstrates potential to be adapted for the use of personality assessment of horses. This method has already been successfully adapted for personality assessment in a wide range of species and would also allow for easier cross-species comparisons.

With reference to Gosling and Vazire (2002) criteria, Criterion One, a novel method of horse personality assessment (i.e. an adaptation of the Stevenson-Hinde *et al.*, 1978 method), could be assessed with relative ease through the use of multiple raters and subsequent correlation analysis (e.g. Morris *et al.*, 2002a; Martin, 2005; Pederson, *et al.* 2005). As such, agreement is usually determined when scores for the same individual animal correlate between different raters. Criterion Two can, however, be more complicated to demonstrate as it requires the prediction of behaviours or real-world outcomes. This has been demonstrated in other species in a variety of ways. In rhesus macaques, for example, Capitanio (1999) was able to test predictions of correlations between personality ratings and behaviours observed in situations independent of the original rating conditions. Furthermore, Capitanio *et al.* (1999) identified significant associations between rhesus macaque personality type and progression of the Simian

Immunodeficiency Virus. In addition, Pederson *et al.* (2005) successfully demonstrated that personality ratings could be used to predict the behaviour of captive chimpanzees. Such tests of prediction would therefore be required in order to evaluate a novel method of horse personality assessment and ensure that Criterion Two can be satisfied. Two ways of approaching this are; to use specific behaviour tests where the relationships between the horses' reactions and specific personality types are predicted and secondly by demonstrating the existence of breed differences in behaviour.

Horse breeds are often described as having 'breed typical behaviours' and such claims are mainly supported by anecdotal evidence from breed enthusiasts, with breed societies often promoting a breed by describing its typical temperament and personality. For example, the highland pony is described as having a "*kindly nature and even temperament*" (Highland Pony Society, 2006) and the Irish draught horse is described as having "*an intelligent and gentle nature and is noted for its docility and sense*" (Irish Draught Horse Society, 2006). In contrast, the Arab has been described as "*spirited, enduring, intelligent, bold, perceptive, sensitive and thoughtful*" (Foster, 2005). Breed differences in horse behaviour have previously been demonstrated by Hausberger *et al*. (2004) who explored the reactions of horses, from 16 breeds, to a bridge test and identified significant breed differences in the length of time taken to cross the 'bridge'. In addition, Hausberger and Muller (2002) found variation in friendly behaviour and reactivity between French saddlebreds, thoroughbreds and Angloarabs. Breed effects have also been identified in the performance of stereotypic behaviours (Luescher *et al.*, 1998; Redbo *et al.*, 1998; Houpt & Kusunose, 2000).

Modern uses of the horse are very diverse and can be both physically and mentally demanding. These uses include competition (e.g. show jumping and racing), army and police work, private leisure use, endurance and the traditional use as draught animals. For

each of these purposes there are specific breeds and types of horses that are selected for their suitability to each of these functions (Clutton-Brock, 1999; Foster, 2005). The primary basis for selection is to select for morphological or aesthetically pleasing characteristics, such as size, strength, colour and speed (Clutton-Brock, 1999; Foster 2005). Nevertheless, desirable behaviours of an individual animal may also be selected for (Hislop, 1992; Houpt & Kusunose, 2000) and may, therefore, have led to breed differences in personality. The demonstration of breed differences in personality that can be explained through the function and ancestry of horse breeds could provide a further way of demonstrating Criterion Two, in that breed differences represent a real-world outcome.

In order to satisfy Criterion Three, (ratings provide a true representation of an individual's personality and are not an artefact of raters' implicit personality theories), Gosling and Vazire (2002) compared personality structures from behavioural studies to those of rating studies. They concluded that similar personality dimensions were being identified by both behavioural and subjective measures of personality. In order for novel assessment methods to be evaluated in terms of Criterion Three it is necessary to compare findings to those of both subjective and behaviour based studies or directly contrast rating data to behavioural observations.

The overall aim of this thesis was, therefore, to develop a method of horse personality assessment and evaluate it for reliability and validity using the three criteria set out by Gosling and Vazire (2002). A reliable assessment method could have several applications and benefits (Mills, 1998), for example, the selection of horses for group-housing situations (Mills, 1998) or the selection of horses for particular disciplines (Visser *et al.*, 2003a). The use of personality assessments in the selection of horses for specific disciplines may also lead to improvements in horse welfare. For instance, horses that suit a

specific discipline would be expected to be less stressed than those with more unsuitable personalities.

1.1 General aims

Aim I

To develop a method for the assessment of horse personality, that meets all three of Gosling and Vazire's (2002) criteria.

- Adapt the Stevenson-Hinde *et al.* (1980) rhesus macaque trait list for the assessment of horse personality.
- In order to meet Gosling and Vazire's (2002) Criterion One personality assessments between raters must be shown to agree, therefore horse personality data will be tested for inter-rater reliability.
- In order to meet Gosling and Vazire's (2002) Criterion Two personality data must be shown to be predictive of behaviour and real-world outcomes. Correlations between horse personality scores and observed behaviour will, therefore, be tested.

Aim II

Explore horse personality data for an underlying structure of personality and compare to both previous horse personality data and that of other species.

- Personality structure to be explored using principal component analysis on horse personality data.
- Comparison of resulting components to identify commonalities with those of other studies. This will further validate the assessment method by demonstrating that personality is being measured and can be linked to Criterion Three.

Aim III

Demonstrate real-world outcomes and satisfy Criterion Two by identifying breed differences in personality.

- Test for significant differences in personality between pure bred horses of eight different horse breeds.

Aim IV

Demonstrate that personality ratings can be used to predict the behaviour of horses in order to satisfy Criterion Two.

- Assess the personality of a selection of horses and then make predictions of how they will behave during three behaviour tests.

2 Literature Review

Personality research in both humans and animals is a rapidly growing area of interest. Sometimes referred to as individual differences, personality has been defined as "*those characteristics of a person that account for consistent patterns of feeling, thinking and behaving*" (Pervin & John, 1997, p.4; Pervin *et al.*, 2005, p.6). This definition emphasises the importance of consistency in the way an individual behaves and infers that these differences be attributed to the individual, as opposed to the environment (Pervin & John, 1997; Pervin *et al.*, 2005, p.6). There are limits to the application of this definition to animals, as the measurement of how animals feel or think is difficult, if not impossible. Animal personality research is therefore focused on the assessment of behaviour in order to demonstrate individual differences and personality in animals.

Personality research on animals has several functions. Primarily it allows personality psychologists to explore the existence and structure of animal personality and to compare to that of humans (Gosling, 2001; Gosling & Vazire, 2002). Animal research can also allow for greater experimental manipulation and multiple generations in a short space of time that would not be possible in human research. This in turn can aid the investigation of the biological basis of personality such as the effects of genetics (Henderson, 1986; Weiss *et al.*, 2000; Dingemanse *et al.*, 2002; Malmkvist & Hansen, 2002; Archer *et al.*, 2003; Mormède, 2005), the effects of personality on the immune system (Capitanio *et al.*, 1999; Capitanio *et al.*, 2004; Cavigelli, 2005) and the biological controls of personality (Boguszewski & Zagrodzka, 2002).

There is also potential for using personality assessment to predict animal behaviour. This would exploit the link between personality and observed behaviour (Byrne & Suomi, 2002;

Pederson *et al.*, 2005) and may become a useful tool in animal management and in the selection of working animals (Mills, 1998; Visser *et al.*, 2003a; Maejima *et al.*, in press). For example, search dogs (Maejima *et al.*, in press), guide dogs (Goddard and Beilharz, 1983, 1984) and performance/sports animals (Visser *et al.*, 2003a). The use of personality assessment in these situations may aid the selection of those animals with the most appropriate personality type for the desired function (Maejima *et al.*, in press).

Despite this, animal personality research is still considered to be a controversial subject. Critics believe that animals can not possess a personality (Kummer *et al.*, 1990; Kennedy, 1992) and consider the use of personality descriptors such as 'caring', 'anxious' and 'calm' as anthropomorphic (Kennedy, 1992; Mitchell & Hamm, 1997) and should be used with caution. Gosling and John (1999) argue against this concern. First they noted that in their review of the animal personality literature, it was demonstrated that independent observers were able to agree about the personality ratings of individuals. Secondly many of the studies they reviewed were carried out using detailed behavioural observations and yielded results comparable to those using rating methods and anthropomorphic terms. Furthermore, it has been noted that the use of anthropomorphic terms is more practical, especially when all one needs to know is whether an animal is agreeable or not (Gosling, 1998), as opposed to a breakdown of its usual behaviours. For example, Hebb (1946) reported on the use of a behavioural recording system in which specific behavioural acts of captive chimpanzees were recorded. These records were meant to provide an objective description of each individual's personality. What resulted, however, were extensive lists of behaviours, which had very little meaning. In contrast, it was found that the use of anthropomorphic descriptors to describe the general peculiarities of the individuals was far more useful than the complicated lists of behaviours, especially for new members of staff. It was therefore concluded that the use of anthropomorphic terms provided "*an intelligible and practical*

guide to behaviour" (Hebb, 1946, p. 88). This is comparable to the idea of a common lexicon that is the foundation of trait theory (John, 1990) (see Section 2.1.3). Furthermore most pet owners are able to describe the personality of their pets and how each individual differs from others of its kind. Breed enthusiasts would also be able to discuss at length the characteristic personality of their favoured breed, such as the kindly nature of a Highland pony (Highland Pony Society, 2006) or the bold, spirited and intelligent Arab (Foster, 2005). Thus these anthropomorphic terms are more easily understood by the layperson and can be more easily applied.

Personality research to date has been carried out on a wide variety of animal species. These include a range of domesticated animals such as cats (*Felis sylvestris catus*) (Feaver *et al.*, 1986), dogs (Goddard & Beilharz, 1984; Goddard & Beilharz, 1986; e.g. Bradshaw *et al.*, 1996; Svartberg, 2002; Strandberg *et al.*, 2005; Diederich & Giffroy, 2006; Christensen *et al.*, in press), cattle (*Bos taurus*) (Kilgour, 1975; Grandin *et al.*, 1995; Kilgour *et al.*, in press) and horses (Le Scolan *et al.*, 1997; Anderson *et al.*, 1999; Visser *et al.*, 2001; Morris *et al.*, 2002a, 2002b; Creighton, 2003; Visser *et al.*, 2003a; Hausberger *et al.*, 2004), as well as wild species (captive and free-range) such as spotted hyena (Gosling, 1998), bottlenosed dolphins (*Tursiops truncatus*) (Kellerman, 1966), octopuses (*Octopus rubescens*) (Mather & Anderson, 1993; Sinn *et al.*, 2001) and several primate species (Stevenson-Hinde & Zunz, 1978; Stevenson-Hinder *et al.*, 1980; Gold & Maple, 1994; Dutton *et al.*, 1997; King & Figueredo, 1997; Lillienfield *et al.*, 1999; Weiss *et al.*, 2000; Martin, 2005; Rouff *et al.*, 2005; Kuhar *et al.*, 2006), with over 64 different species having been studied (for reviews see; Gosling and John, 1999; Gosling, 2001; Gosling & Vazire, 2002).

The purpose of this review is to provide an understanding and critique of both the historic and current approaches used to assess animal personality and to explore the current evidence and our ability to assess it accurately. Some reference to human personality research will be required in order to further understand the origins of animal personality research. In addition there will be particular focus on horse personality research with respect to the methods used, current findings and the potential applications of a reliable assessment method within the equine industry.

2.1 Main approaches to personality research

There are a wide variety of approaches used in the assessment of personality. Much of the early work began with psychoanalysis and Freudian theories (e.g. Freud, 1961) and Plutchik's (1960) theory of emotion. These appear to have fallen out of favour during the late 1900s and were replaced by trait theory and simple rating approaches (e.g. Eysenck, 1991). Additionally, in terms of animal personality research, behavioural measures of personality have also been used extensively (e.g. Goddard, 1986; Anderson *et al.*, 1999; Archer *et al.,* 2003; Diederich and Giffroy, 2006). This section of the review discusses each of these key approaches, highlighting the advantages as well as the limitations of each, especially with respect to the evaluation of animal personality assessment and research.

2.1.1 Freudian theories

The psychoanalytic approach, founded by Freud (1961), was one of the earliest personality theories. Psychodynamic theories are defined by Gross *et al.* (2000) as "*those which focus on the active forces within the personality that motivate behaviour, and the inner causes of behaviour, in particular the unconscious conflict between the various personality structures*". Key theories in psychoanalysis include the topographic, developmental and structural models (Pervin and John, 1997). Freud's final, and perhaps most familiar, model of personality was the structural model (Freud, 1961), which categorised mental processes by their functions or purposes. The model comprised the id, ego and superego and involved conflict between these three unconscious drives (Western, 1990; Pervin and John, 1997). The id represents the drive of an individual and operates on the basis of primary thought. It seeks pleasure and can often be unrealistic. In contrast, the superego provides the individual with morals and ethics. It determines the ideals humans strive for and the

punishment (i.e. guilt) expected when the ethical code is broken. The ego, however, seeks reality and is the controller of the id and the superego, and strives to find a balance between the two. The ego aims to provide the id with satisfaction of its demands, whilst satisfying the morals, ethics and rules of the superego. The ego has the role of trying to separate wish from fantasy and thinks more in terms of the future as opposed to allowing the id everything when it wants. Freud and other psychoanalytic researchers developed several theories with different emphasis on the id, ego or superego (Western, 1990). Although popular in the early 1900s, psychoanalytic theory has since lost much of its support. One reason for this is the lack of testable theories and the limited empirical evidence (Eysenck, 1991). As a result of this and the difficulty of applying Freudian theories to animals, Freud's psychoanalytic theories have not been applied to animal personality research.

2.1.2 Theory of emotion

Plutchik's (1960; 1965) multifactor-analytic theory of emotion is based on psychoanalytic theory and describes behaviours in terms of bipolar tendencies towards action. These are; *Moving Toward* (destruction) versus *Moving Away* (protection); *Taking In* (incorporation) versus *Expelling* (rejection); *Possessing* (reproduction) versus *Losing* (deprivation) and *Moving* (exploration) versus *Stopping* (orientation). "*These eight patterns of reaction involve the whole organism and have introspective, behavioural and physiological aspects*" (Plutchik, 1965, p.296). These patterns were termed primary emotions, and all other emotions were described as a combination of these (Plutchik, 1965). Plutchik (1960; 1965) linked emotion to personality by proposing that the formation of personality traits was related to the development of mixed emotions and that all personality traits implied emotional components at some level of conflict (Plutchik, 1960; 1965). The assessment of personality was carried out using the Emotional Profile Index (EPI), which was

constructed using the eight primary emotions. Plutchik's theory of emotion has been successfully applied to both humans (e.g. Kellerman & Plutchik, 1968; e.g. Platman *et al.*, 1969) and animals (Kellerman, 1966; Buirski *et al.*, 1973; Buirski *et al.*, 1978).

Animal studies based on Plutchik's (1960; 1965) theory of emotion initially worked on bottlenosed dolphins (Kellerman, 1966) and later on several primate species (Buirski *et al.*, 1973; Martau *et al.*, 1985; Buirski & Plutchik, 1991). Kellerman's (1966) dolphin study utilised an EPI based on Plutchik's (1960; 1965) theory of emotion. The dolphin EPI was based on eleven behaviourally defined adjectives, paired into all possible combinations, totalling 52 pairs. Raters chose the adjective out of the pair that best described the dolphin in question. Each adjective was defined by a combination of two or more of the eight primary emotions. The result was eight scores for each individual, representing the strength of each primary emotion. For example, Kellerman (1966) proposed that if a dolphin was considered to be aggressive then the underlying dimensions involved might be anger and the anticipation of an event. The behaviour would therefore be analysed in terms of its emotional dimensions, i.e. exploration and destruction. Using the dolphin EPI Kellerman (1966) identified individual differences between the three dolphins studied. Two of the dolphins were rated as being similar to each other but different to the third dolphin. These data provided early evidence of our ability to measure emotion (personality) and individual differences in animals.

Buirski *et al.* (1973) carried out a similar study on a troop of seven free-ranging olive baboons (*Papio anubis*). The baboon EPI was developed using 12 descriptive terms in the same way as Kellerman's (1966) dolphin EPI. The troop was observed for a total of 35 hours over a three-week period by three observers who were later required to complete the EPI for each baboon observed. The resulting emotion scores were then correlated against

dominance and grooming indices in order to explore any possible relationships. Significant correlations were found between 'time being groomed' and four of the primary emotions, *Protection*, *Deprivation*, *Rejection* and *Destruction*. A relationship was also found between sociability ratings and dominance, such that the more dominant animals were found to be less sociable and spent more time being aggressive, when compared to subordinate animals. These correlations suggested a relationship between personality and observed behaviour and inferred that the raters were providing genuine and accurate assessments of the individuals.

The use of Plutchik's (1960; 1965) theory of emotion and the EPI successfully identified individual differences in humans and in a range of animals. The theory of emotion is, however, subject to the same criticisms as other psychoanalytic theories. The conceptual nature of the theory makes it difficult to test, resulting in limited empirical evidence (Eysenck, 1991). It is further criticised for the fixed nature of the theory and the highly subjective manner in which the primary emotions are selected and paired (Eysenck, 1991). Use of the EPI to assess personality in animals has also been questioned due to its dependence on a theory originally developed for humans (Stevenson-Hinde & Zunz, 1978; Stevenson-Hinde *et al.*, 1980). For example, Stevenson-Hinde and Zunz (1978) and Stevenson-Hinde *et al.* (1980) comment on how the EPI assumes that the same components (emotions) apply to both humans and other animals. They suggest this as a possible weakness of the EPI with respect to animal personality assessment, as it assumed that humans and animals would have the same personality structure. As a result of these criticisms there has been a reduction in the popularity of this theory in both human and animal personality research. Attention has instead turned towards the development of the more flexible approach of trait theory.

2.1.3 Trait theory

In the 1930s interest in psychoanalytic theory began to be replaced by trait theory, which was seen as a more extensive and flexible approach to the assessment of personality in both humans and animals. Trait theory is based on the lexical hypothesis that the most socially relevant and important personality characteristics are encoded in the natural language (John, 1990; Pervin *et al.*, 2005) and these are often referred to as traits. Traits are defined as "*generalised and personalised determining tendencies – consistent and stable modes of an individual's adjustment to his environment*" (Allport & Odbert, 1936, p. 26). The assessment of personality using traits and subsequent analysis using multivariate statistics allows for a detailed exploration of personality taxonomy (i.e. the grouping of similar or related personality traits into specific categories or dimensions). This is in contrast to the fixed and subjective theories of psychoanalysis and the theory of emotion.

Early work on trait theory focused on the development of trait lists from lexicons and the natural language. Researchers selected those adjectives (traits) that best described elements of personality. Works by Allport and Odbert (1936), Cattell (1943; 1945; 1947) and Norman (1963) were highly influential in the creation of trait lists and the initial exploration of personality taxonomy. Trait lists could then be used to assess personality of individuals by rating them usually with five or seven point Likert-type scales (Coolican, 2004). The resulting data could then be analysed using multivariate statistical techniques that reduced the data into personality dimensions or factors. For example cluster analysis, factor analysis (FA) and principal component analysis (PCA).

Factor analysis is a statistical procedure that is used to identify whether a factor structure underlies correlations between multiple variables (Brace *et al.*, 2003). Principal component analysis is similar but uses different extraction methods which maximise the variance

explained by the resulting model (Brace *et al.*, 2003). Principal component analysis and FA group correlated items into components or factors that are unrelated to each other. The combination of variables that account for most of the variance are identified and grouped together to form the first component. The second component becomes the combination that accounts for the second greatest amount of variation, and so on, until all the variance is accounted for (for more information see; Kline, 1993a,1993b; Tabachnick & Fidell, 2001; Brace *et al.*, 2003).

In personality research, the resulting factors or components are interpreted as being dimensions of personality and these are similar in concept to Plutchik's (1965) primary emotions. The major difference is that the components/factors are not predetermined by human selection; they are instead reliant on inputted data and statistical analysis. In addition PCA or FA components are not fixed and may vary between species, in both number and structure. Furthermore, personality assessment may involve many traits that cover a wide range of personality descriptors. Data reduction using FA or PCA makes interpretation of personality data simpler, as it avoids trying to interpret the results from many different variables. The number and composition of personality factors are, however, major areas of discussion within trait based personality research.

Cattell (1943), for example, used factor and cluster analyses to develop a taxonomic structure of human personality. His research resulted in 12 personality factors (Cattell, 1945), and further encouraged researchers to explore the structure of personality. Later, Fiske (1949) adapted Cattell's (1943; 1945) trait lists and used them to obtain trait ratings of a sample of 128 clinical trainees. The resulting data suggested a taxonomic structure of five factors. Tupes and Christal's (1961; cited in John, 1990) study on a range of different samples (e.g. airmen and first year graduate students) also showed a strong and recurrent

five-factor structure. These factors were labelled as *Surgency*, *Agreeableness*, *Dependability*, *Emotional Stability* and *Culture*. Similar five-factor structures have also been identified by Norman (1963), Borgatta (1964) and Digman and Takemoto-Chock (1981).

The stability of the five-factor model has been demonstrated in several human rating studies. Using an abbreviated version of Norman's (1967) list, Goldberg (1990) assessed two samples of peer ratings and self-ratings. The data showed consistency in both the occurrence and the construction of the five factors. Factor five, however, was found to represent a more intellectual basis than that of culture as suggested by previous studies, including those of Norman (1963) and Tupes and Christal (1961; cited in John, 1990). This supported previous findings of McCrae and Costa (1985; 1987) who described factor five as *Openness to experience,* also known as *Openness*.

The five-factor model (*Neuroticism, Extraversion, Conscientiousness, Agreeableness,* and *Openness*) is now widely accepted and has been used to assess several aspects of personality research including personality stability in adulthood (reviewed by McCrae, 2002), heritability (Loehlin *et al.*, 1998) and cross-cultural comparisons (McCrae & Costa, 2004). Although there is still some discussion amongst personality researchers as to the full acceptance of the five-factor structure (also known as the Big Five), it has become a widely used taxonomy of personality. As a result, the assessment method developed by Costa and McCrae (1992a) has also become favourable. This method of assessment employs a 60-item questionnaire which measures each of the Big Five using twelve items per factor and a five-point rating scale. It has been translated into several different languages and is one of the most widely used measures of human personality (McCrae & Costa, 2004). Although widely supported within personality research, it should, however, be noted that

not all personality researchers accept or utilise the five-factor taxonomy of personality (e.g. Eysenck, 1991, 1992; Gana and Trouillet, 2003; Gillespie, *et al.*, 2003).

In summary, personality research on humans has gone through a significant amount of development in the last century. The original theories of psychoanalysis, although still in use, have decreased in frequency and have for the most part, been overtaken by more trait-based methods of research that seek to define the taxonomy of personality. The concept of personality is now widely accepted within psychology and is fuelling much research into a wide range of subject areas. Studies include the exploration of different personality types associated with the development of disease (Capitanio *et al.*, 1999; Marušič & Eysenck, 2001; Garssen, 2004), links between personality and neurological systems (reviewed by Rowe, 1995), performance during sport (Egloff & Gruhn, 1996), intelligence (Moutafi *et al.*, 2005) and training performance (Dean *et al.*, in press).

Trait rating approach in animals

The application of trait theory to animals has become common in animal personality research. In animals, this approach involves the rating of an animal against a list of behaviourally defined adjectives or traits. Observers, usually the regular handlers of the animals, rate individuals on a five- or seven-point scale for each of the traits, and assess the individuals' general behaviour tendencies during the time they have known them. This method of assessment can be compared to a peer rating system used with humans and assesses the public or perceived personality of the individual.

Being restricted to an observer-rating system could be seen as a limitation of animal personality research. Research in humans has, however, found that observer-observer agreement tends to be high (Funder & Dobroth, 1987; Funder *et al.*, 1995) and can be

higher than that of self-observer agreement (Gosling, 1998). This is thought to be as a result of the 'self' enhancing those qualities that it sees as more desirable (Gosling, 1998), i.e. a difficulty in being critical of oneself. Additionally human studies have identified an acquaintance effect. Such that individuals who know the subject being rated, rate them more accurately than would a stranger (Funder *et al.*, 1995), as a result of being able to observe the subject over a longer period of time. It therefore seems logical that regular handlers or observers of animals would be able to accurately rate the personality of individuals. The use of a subjective rating method should, therefore, be a reliable and accurate method of assessing animal personality.

This approach was pioneered for the assessment animal personality assessment by Stevenson-Hinde and Zunz (1978) and Stevenson-Hinde *et al.* (1980) who adapted it for the personality assessment of 45 captive rhesus macaques (*Macaca mulatta*). The original list of 33 behaviourally defined adjectives (traits) was derived from terms regularly used by people working with the macaques to describe macaque behaviour. Three handlers later rated each monkey on a seven-point rating scale, on each of the behaviourally defined adjectives. Where one represented extreme antithesis and seven represented extreme manifestation. The questionnaire was later revised and resulted in a 25-item trait list (Stevenson-Hinde *et al.*, 1980). In addition, the personality of each of the 45 rhesus macaques was assessed over four successive years. Principal component analysis of the data for the first two years extracted two components, which were shown to be stable over the four years. After the addition of new adjectives in the third year, a three-component structure was extracted, and was shown to be stable for the third and fourth years, explaining 66% (third year) and 69% (fourth year) of the variability. Through examination of the major contributing traits on each component, the components were described as *Confident-Fearful*, *Active-Slow* and *Sociable-Solitary*. Personality stability was also

demonstrated once the rhesus macaques reached maturity, having shown more unstable personalities at juvenile stages.

This pioneering approach to animal personality demonstrated that the trait rating method could be applied to animals and provided a relatively simple and effective assessment method that could be used by both researcher and layperson. These studies acted as a catalyst for further animal personality research using trait theory, resulting in the trait list being adapted for a variety of different species. These have included: cats (Feaver *et al.*, 1986); gorillas (*Gorilla gorilla*) (Gold & Maple, 1994); spotted hyenas (Gosling, 1998); cheetahs (*Acinonyx jubatus*) (Wielebnowski, 1999); chimpanzees (Martin, 2005) and pig-tailed macaques (*Macaca nemestrina*) (Caine *et al.*, 1983). Furthermore, researchers have sought to develop their own assessment methods and trait lists which incorporated trait theory (McGuire *et al.*, 1994; Dutton *et al.*, 1997).

McGuire *et al.* (1994) created their own adjective list to assess the personality of 97 vervet monkeys (*Cercopithecus aethiops sabaeus*). Seven experienced observers created a list of twelve personality constructs (traits) using specific selection criteria. A construct had to be applicable to at least half of the test animals and not be readily codeable using traditional behavioural recording techniques. The twelve constructs and an additional five ethologically coded behaviours were then used to assess personality. The resulting list was similar to that of Stevenson-Hinde *et al.* (1980), but was rated using a five-point rating scale. When rating the animals, observers were instructed to assign the average score (3) to 40% of the individuals, intermediate scores (2 and 4) to 20% of the individuals and each of the extreme scores (1 and 5) to 10%. Principal component analysis extracted three personality components which accounted for 70% of the variance. These were described as *Socially Competent*, *Playful/Curious* and *Opportunistic*. The rating method and

prescriptive guidelines used by McGuire *et al.* (1994) are, however, quite restrictive. The instruction to raters that they should meet specific quotas for each rating category seems to conflict with the concept of personality. Although it is natural to compare the behaviour of an individual to that of its conspecifics, by limiting the number of individuals allowed in each category may result in the inaccurate rating of borderline individuals, and assumes a normal distribution of personality types within the sample population. This may be further exacerbated by the limitations of a five-point rating scale. The choice of rating system should therefore be considered carefully during the development of assessment methods.

Although commonly used, the Stevenson-Hinde and Zunz (1978) approach has also been criticised. For example, Dutton *et al.* (1997) remarked on its dependence on an *ad hoc* classification of behaviours. Those methods that were based on human assessment techniques (e.g. Kellerman, 1966; Buirski *et al.*, 1973) were also criticised by Dutton *et al.* (1997) for the assumption that animals would conform to the same rating systems and structures as humans. Dutton *et al.* (1997) proposed a novel assessment method that was based on the constructivist approach of Kelly (1955) and used a repertory grid to generate personality constructs. Raters were essentially allowed to select their own set of constructs (traits) using the grid and used these to rate the subjects, a troop of captive chimpanzees. Each rater's data were individually analysed using PCA and the resulting personality dimensions were compared across raters for similarity. Four relatively stable personality factors were identified: *Dominance, Sociability, Machiavellianism* and *Anxiety*. They were not, however, found in all seven raters' solutions, but *Dominance* and *Sociability* were found in six out of the seven. The construction of these factors appeared similar to those found in other primate studies, e.g. rhesus macaques (Chamove *et al.*, 1972; Stevenson-Hinde & Zunz, 1978; Stevenson-Hinde *et al.*, 1980) and vervet monkeys (McGuire *et al.*, 1994). Dutton *et al.* (1997) also found similarities with the three-factor structure of human

personality as identified by Eysenck (1952). Although similar in concept to the Stevenson-Hinde and Zunz (1978) approach of personality assessment, Dutton *et al.* (1997) allowed for greater flexibility in the selection of traits. This flexibility may, however, result in difficulties in assessing rater agreement and may reduce the consistency in trait interpretation by raters. This method should therefore be approached with some caution.

In contrast, King and Figueredo (1997) adapted human trait lists developed by Norman (1963) and Goldberg (1990), to explore chimpanzee personality taxonomy and compare it to that of humans. A total of 43 adjectives were selected from the human lists and were intended to be representative of each of the five factors of human personality. Principal component analysis of ratings by 53 raters of 100 chimpanzees (mean of 4.05 raters per chimpanzee) resulted in a six-factor taxonomy. The first and strongest factor was interpreted as a *Dominance* factor, with the other five thought to be comparable to the 'Big Five' of humans, and were described as: S*urgency; Dependability/Conscientiousness; Agreeableness; Emotional/Neuroticism* and *Openness*. This similarity in personality taxonomy is not surprising considering the close phylogenetic relationship of chimpanzees to humans (Sibley & Ahlquist, 1984). To date this six-factor structure is the closest match to the human Big Five to have been identified in non-human species. Although a five-factor structure has been identified in chimpanzees by Martin (2005), the component structure did not closely fit that of the Big Five.

Trait theory and equine personality assessment

No studies to date, have applied the Stevenson-Hinde *et al.* (1980) adjective list to equines, although other rating systems have been employed. An early example of equine personality research is French's (1993) study on the temperament of donkeys (*Equus asinus*). By adapting rating methods originally developed by Stevenson-Hinde and Zunz (1978),

French (1993) rated re-homed donkeys using a continuous line-scale between paired adjectives. This involved a measured line of fixed length upon which the raters marked a cross at the point which best described the individual's personality with respect to the traits being rated. The score was then given as the distance (millimetres) from the left hand side of the line and provided the observer with a greater choice of score. The measurement of the line-scales can, however, be time consuming, and will incur some measurement error. The sixteen-paired adjectives used to form the questionnaire were; staid-playful, shy-outgoing, calm-nervous, gentle-rough, obliging-wilful, friendly-spiteful, content-agitated and handling easy-difficult. These terms were thought of as being easily used to describe personality in both humans and animals. Principal component analysis on data from 45 donkeys, extracted two components described as *Vivacity* and *Obduracy*. This rating method was later adapted by Anderson *et al*. (1999) who used twenty paired adjectives and a five-point rating system to assess the personality of 103 horses. Rater-reliability was, however, found to be quite low and will be discussed later in more detail (Section 2.2.1). In addition to the personality ratings, Anderson *et al*. (1999) measured blood plasma cortisol concentrations and carried out novel object tests during which the behaviour of the horses was recorded. Relationships between these measures and the assessed personality of the horses suggested a relationship but did not provide conclusive evidence as to the validity of the trait based assessment method.

More recently Momozawa *et al.* (2003) developed a questionnaire consisting of eight, non-defined, temperament adjectives, each rated on a five-point rating scale. Regular handlers were used to assess 86 horses and the mean scores were calculated and used for further analysis. Factor analysis on these data extracted three factors and accounted for 84% of the variance. The factors were described as *Anxiety*, *Novelty Seeking* and *Understanding*. The lack of any behavioural definition, or explanation of each trait, was identified by

Momozawa *et al.* (2003) as a weakness of the study. As a result, Momozawa *et al.* (2005) developed the questionnaire further by expanding the list to 20 and behaviourally defined each item. This was thought to minimise the variation among raters due to linguistic misinterpretation. The rating scale was also increased to a nine-point rating scale. The study took place over two years and rated 69 horses in the first year and 70 horses in the second year. Factor analysis was then used to explore the personality taxonomy of the horses and factor structure was compared between the two years. Five factors were extracted in both years and accounted for 71.4% and 75.5% variability. Of these factors, consistency in structure between years was found in four of the five factors. These were termed, *Anxiety*, *Trainability*, *Affability* and *Gate entrance* (so called, as this was the only trait that significantly loaded on to this factor).

In addition to the trait-based assessments, Momozawa *et al.* (2003) carried out a behavioural test in order to assess the validity of their method. The horses were released, for five minutes, into an unfamiliar indoor arena (7 m x 12.5 m x 3 m) at the centre of which hung two slowly revolving balloons. The responses of the horses were assessed by measuring alterations in hear rate (HR) and behavioural observations (recorded remotely) during the five minute experimental period. Spearman rank order correlations were carried out between the balloon test measurements and the questionnaire/rating data. Change in heart rate (i.e. between experimental and control conditions) (r_s = 0.312, P<0.01) and frequency of defecation (r_s = 0.26, P<0.05) were both found to significantly correlate with *Anxiety*. These correlations indicated that the questionnaire was a valid measure of horse personality although such links with observed behaviour were limited.

The trait rating method used by Momozawa *et al.* (2005) indicated that personality could be reliably assessed in horses and produced stable personality components across years.

Few of the traits used in the 20-item questionnaire are, however, comparable to those used in trait rating studies on other species, thus making cross-species comparisons difficult. Similarly, the use of human based rating systems limits cross-species comparisons of personality structure.

In contrast to most trait rating studies on animals, Morris *et al.* (2002a) used and McCrae's (1992a) Revised NEO Personality Inventory and NEO Five Factor Index (NEO-PI-FFI) to measure horse personality and test judge reliability. The NEO-PI-FFI is a popular rating tool used to assess human personality. The study used the original 60 items from the NEO-PI-FFI, with each converted to the third person for ease of use by horse handlers. Each of the 'Big Five' factors was represented by twelve items, which were structured as statements. For example, on the *Neuroticism* factor items included "*he/she worries a lot*" and "*he/she is likely to be discouraged and give up*" (Morris *et al.*, 2002, p. 73). Each observer was asked to assess study horses on each item by using labelled analogue scales, which measured 32 mm (similar to the continuous line-scale used by French, 1999). For example, a score of 32 mm showed strong agreement and a score of 0 mm represented strong disagreement to a statement. The study used ten horses with each horse being assessed by nine judges. The main aim of the study was to explore inter-rater reliability as opposed to personality taxonomy. High inter-rater reliabilities were found and will be discussed in more detail in Section 2.2.1 of this review. It was demonstrated that the *Neuroticism* and *Extraversion* scales were easiest to apply to horses whereas *Openness* and *Conscientiousness* were more difficult to apply. This may have been as a result of the relevance of some of the items with respect to horse personality. On the *Openness* factor, the following items for example, would be very difficult to assess in terms of horse personality and behaviour; "*has day dreams but does not like day dreaming*", "*spends time speculating about the nature of the universe*" and "*likes poetry*" (Morris *et al.*, 2002a, p.

73). As well as being difficult to assess, statements such as these have very little relevance to the practical applications of horse personality. Although many of the items used in the NEO-PI-FFI are easily transferable to horses and other animals (e.g. those items measuring *Neuroticism* and *Extraversion*), others are far more difficult to apply and, in some cases, illogical and irrelevant. Furthermore the validity of a human rating questionnaire to assess the personality of horses is questionable as it assumes that humans and horses will share the same personality structure. Dependence on human personality structures and theory have already been criticised by Eysenck (1991), Stevenson-Hinde *et al.* (1980) and Stevenson-Hinde and Zunz (1978). The small sample size used in this study also makes accurate interpretation of the data difficult. A horse personality assessment tool based on the Stevenson-Hinde *et al.* (1980) trait list would, however, allow for more flexibility and easier cross-species comparison of personality structure. Such comparisons may then provide further understanding of the evolution and development of personality.

The few studies that have approached the use of trait rating questionnaires for the assessment of horse personality have indicated that trait rating can be a reliable method of personality assessment of horses. The validity of these assessments, however, is yet to be fully demonstrated. Nevertheless, many studies of individual differences in horses incorporate the use of behavioural tests and observations instead of rating questionnaires. The benefits and limitations of these approaches with respect to horse personality assessment are therefore discussed later in Section 2.3.

In summary, the application of trait theory for the assessment of both human and animal personality has become common. The main benefits of using a trait rating approach is that it is quick and easy to apply and the use of multivariate statistics allows for the exploration of personality taxonomy. Use of trait theory in animal personality research allows for a

flexible analysis of personality structure that does not rely on human taxonomic theories, unlike the EPI approach developed by Plutchik (1960; 1965). Adaptation of the Stevenson-Hinde (1980) trait list for the assessment of a variety of species, allows for some cross species comparisons (Gosling & John, 1999; Gosling, 2001; Gosling *et al.*, 2003), whilst still allowing for flexibility in the choice of traits used. This, therefore, ensures that each trait list is relevant to the species being assessed. This is more difficult in those approaches that incorporate human rating lists and fixed taxonomies (King & Figueredo, 1997; Morris *et al.*, 2002a, 2002b), as many traits that may be easily rated or recognised in humans may not be relevant to non-humans. Trait rating approaches should therefore be developed carefully and with great consideration for the relevance of traits to the behaviour of the species being assessed. Furthermore, it should be assumed, until evidence can be provided to the contrary, that animal personality taxonomy does not match that of human taxonomy.

2.1.4 Behavioural Assessment of Personality

The use of behavioural measures and tests is a common approach for personality assessment in animal research (e.g. Goddard & Beilharz, 1986; Anderson *et al.*, 1999; Archer *et al.,* 2003; Diederich and Giffroy, 2006). It is considered to be a more objective approach than trait rating as behaviours are quantified either by direct measurement or transformed into ordinal data by categorising different levels of behaviour (e.g. agonistic behaviours may be ranked from one – low to five - high) (Manteca and Deag, 1993). Such assessments are thought to be less open to anthropomorphic projections (Gosling & Vazire, 2002) than subjective assessments of animal personality. Behaviour based studies are also utilised for testing the consistency of behaviours in specific situations (Manteca and Deag, 1993), an integral part of personality that is difficult to test in trait-based research. The use of behaviour tests are, however, facilitated in the assessment of

temperament as well as personality, potentially leading to some confusion over the constructs being measured.

Temperament is defined as *"A characteristic response style to novel stimuli or challenging situations that is largely independent of immediate social or volitional influences..."* (Clarke & Boinski, 1995, p.106). In contrast to personality, temperament is generally seen as a characteristic response or reaction to a certain or specific situation with consistency over time and across situations (Zuckerman, 1991b; Visser *et al.*, 2001). Both terms clearly emphasise the importance of consistency and are therefore very similar. It is often noted, however, that the key difference between personality and temperament is that temperament is independent of social influences (Clarke & Boinski, 1995) and has a greater dependence on environmental effects. It would seem, however, that as these terms are often used interchangeably and are not always clearly defined in the literature, that they are essentially the same concept. It is also possible that some authors are reluctant to use the term personality with respect to animals, as it is something that may be deemed by some, to be a purely human quality. The term temperament has, instead, often been confined to the description of animals and very young children. In terms of this thesis, the term temperament has been accepted as forming an integral part of personality. Thus those studies that claim to assess temperament have been included within the review of the literature. It should be noted, however, that throughout this review, the term temperament has been used when the original authors had used this term. Such studies have often focused on the use of behavioural tests in order to measure individual differences in behaviour.

At a basic level, simple field observations can be used to provide information on individual differences in a group of animals. For example, Maestripieri (2000) observed the

scratching frequencies of five adult female rhesus macaques living in captive social groups. Individual differences were identified in scratching frequency and were consistent across two consecutive birthing seasons, but tended to be higher during the birthing seasons than the mating season. It was suggested that the presence of infants was associated with increased levels of emotionality and the measurement of scratching frequency provided a non-invasive assessment of emotionality in rhesus macaques. This approach to assessment of individual differences is however, very restricted in what can be interpreted from it. Individual differences may occur across several behaviours and situations and measurement should therefore not be restricted to specific behaviour patterns.

Behaviour tests are also utilised in the assessment of individual differences in animals. Reactions of individuals to the tests are recorded and are compared to those of other individuals. For example, Wolff *et al.* (1997) used an arena test in addition to a novel handling and novel object test to assess the personality (emotionality) of 1-3 year old horses of the same breed (French saddlebreds). The tests aimed to assess the gregariousness and fearfulness of the horses. Individual differences were identified in the responses of horses to each test, despite the authors controlling for breed, age and management effects, thus demonstrating real variation in behaviour tendencies between individuals. The study also identified some similarities in behaviour between paternal half-siblings, and suggests some genetic links to personality.

Behavioural tests similar to these have also been applied in other studies of horse behaviour (Le Scolan *et al.*, 1997; Visser *et al.*, 2001; Seaman *et al.*, 2002; Visser *et al.*, 2002; Hausberger *et al.*, 2004; McCall *et al.*, 2006). Such studies have identified individual differences in behavioural tendencies, but are restricted to the element of personality that

they measure, for example reactivity (Hausberger *et al.*, 2004). The use of open-field and novel object tests has also demonstrated individual differences in other species, for example, pigs (Spoolder *et al.*, 1996; Ruis *et al.*, 2000), cattle (Boissy & Bouissou, 1995; Muller & Schrader, 2005; Kilgour *et al.*, in press), passerines (Verbeek *et al.*, 1994; Dingemanse *et al.*, 2002, 2003; Drent *et al.*, 2002; van Oers *et al.*, 2003; Carere *et al.*, 2005; Mettke-Hofmann *et al.*, 2005) and sheep (Romeyer & Bouissou, 1992; McBride & Wolf, in press). Although capable of identifying individual differences in behaviour, these tests are restricted in which components of personality they measure, for example reactivity or sociability. Furthermore, they can be quite time consuming and demanding of resources.

The measurement of individual differences is only a small part of measuring personality. These differences must be demonstrated as being consistent over time and across situations (Pervin and John, 1997). Some studies have attempted to measure consistency in behaviour of individuals with varying degrees of success. Visser *et al.* (2001) used novel object and handling tests to assess the personality of 41 Dutch warm-blood horses. The horses were tested at 9, 10, 21 and 22 months of age and a number of behavioural variables were recorded. Behavioural response variables were found to correlate within years but few variables showed consistency over years. It was concluded that the behavioural tests used could reliably assess personality and temperamental traits in horses, but that the long-term consistency of response needed to be tested further. Visser and associates have since used these and similar methods of behavioural measurement for exploring links between heart rate variability and behaviour (Visser *et al.*, 2002) and predicting show jumping performance in young horses (Visser *et al.*, 2003a). Consistency in horse behaviour has also been demonstrated by Seaman *et al.* (2002) in reactions of 33 horses to an arena test. These results were not, however, shown to be predictive of reactions in the other behaviour

tests performed during the study (person tests, object test, startle test and reaction to being returned to conspecifics). Although there is some evidence of short-term consistency in the behaviour response of horses, it appears that more research is required to demonstrate long-term and cross-situational consistency.

Similar problems with consistency have also been identified in pigs. Spoolder *et al.* (1996) assessed the consistency of behavioural responses in pigs using four different behavioural tests; 1) open field with novel object; 2) individual access to food for 15 minutes after 20 hours food deprivation; 3) competition for food after deprivation and 4) general activity and feeding behaviour in a group for 24 hours. Tests 1 and 2 were repeated four and three times over a two and one week period respectively. The results demonstrated some short-term consistency in behaviour but could not confirm strong inter-situation correlations. In contrast use of an arena test by McBride and Wolf (in press) to assess individual differences of sheep demonstrated good consistency between three repetitions of the test. From these examples it is clear that although these particular types of behaviour tests are capable of demonstrating individual differences in behaviour, they are less reliable at producing consistent results over time or across situations.

Other behavioural tests have been facilitated in the assessment of personality in animals. For example, McCann *et al.* (1988) used a chute test to assess horse personality. Horses were observed during entry and exit of a chute and during the process of fitting horses with identification collars. Four observers rated the horses, using a four-point scale, on their reactions to each phase (on scale from highly nervous to quiet). Inter-rater agreement was high, especially for horses scored as normal, compared to those rated as highly nervous. Furthermore behaviour ratings were significantly correlated with heart rate whilst the

animals were in the chute. These results indicate that this method is capable of providing reliable measurements of individual differences in horses.

The assessment of dog personality has been approached through both the use of rating methods (Serpell & Hsu, 2001) and behavioural tests (Cattell, 1973; Goddard & Beilharz, 1984; Goddard & Beilharz, 1986; Svartberg, 2002; Svartberg & Forkman, 2002; Strandberg *et al.*, 2005; Svartberg, 2005; Svartberg *et al.*, 2005; Svartberg, 2006), with greater emphasis on the latter (for reviews see; Jones & Gosling, 2005; Diederich & Giffroy, 2006). For example, the Swedish Working Dog Association (SWDA) developed a group of behavioural tests for the purpose of assessing personality in dogs. The Dog Mentality Assessment (DMA) was developed mainly as a tool for the breeding of working dogs, but has now been adapted for use with other breeds.

During the DMA, the dogs are exposed to several different situations and their reactions described by trained and official observers, using a standardised score sheet. Emphasis is placed on an objective and neutral description of the behaviours and inter-observer reliability is regularly tested during observer training. The DMA consists of ten separate subtests, which are performed outdoors and in a specific order. The subtests are as follows; social contact, play 1, chase, passive situation, distance play, sudden appearance, metallic noise, 'ghosts', play 2 and gunshot. For detailed descriptions of the subtests involved with the DMA see Svartberg and Forkman (2002).

Using the vast amount of data available via the SWDA Svartberg and Forkman (2002) carried out a factor analysis on data from 1175 dogs (47 breeds each represented by 25 randomly selected dogs) on 33 behavioural variables recorded during the DMA. The factor analysis extracted five factors, which were labelled *Playfulness*, *Curiosity/Fearlessness*,

Chase-proneness, *Sociability* and *Aggressiveness*. Factor analysis of within breed group data demonstrated that these factors were consistent throughout the different breeds of dog and were representative of dog personality structure. This study demonstrates that the use of a wide variety of behaviour tests can provide a detailed picture of animal personality. In order to do this, however, it is necessary to use tests that are thought to assess different elements of personality. The DMA appears to successfully assess a variety of elements of dog personality, including social and play behaviours, aggression, fear and curiosity, all of which are of relevance to the owners and handlers of these dogs. Data from the SWDA have also been used to assess consistency of dog personality (Svartberg *et al.*, 2005), links between DMA scores and the everyday behaviour of dogs (Svartberg, 2005) and breed typical personality of dogs (Svartberg, 2006).

The use of multivariate analysis techniques similar to those used by Svartberg and Forkman (2002) are becoming increasingly common in the analysis of behavioural data. Such analysis allows the researcher to group similar behavioural reactions and potentially explore the taxonomy of personality. This approach has been used to assess behavioural data of dogs (Cattell, 1973; Svartberg & Forkman, 2002; Svartberg *et al.*, 2005), rhesus macaques (Chamove *et al.*, 1972), octopuses (Mather & Anderson, 1993; Sinn *et al.*, 2001), horses (Visser *et al.*, 2001; Visser *et al.*, 2003a), cherry salmon (*Oncorhyncus masou macrostomus*) (Iguchi *et al.*, 2001) and sheep (Ovines) (McBride & Wolf, in press). These studies provide further information regarding the taxonomy of personality that can then be compared to similar analyses on trait-rating data.

The assessment of personality using behavioural tests can also be carried out in conjunction with trait rating. Such an approach aims to demonstrate validity of the methods employed. For example, Le Scolan *et al.* (1997) used behavioural tests to assess the

temperament of 72 horses in a variety of situations. The study compared the observed behaviours during behavioural tests (novel object, arena test, bridge test, instrumental task and spatial task) to ratings of temperament traits such as 'fearful when ridden' and 'socially dependent in an unknown surrounding'. The study aimed to explore the traits of fearfulness, gregariousness, nervousness and learning and memory. These traits were all familiar terms to both riders and handlers. The assessment of the behavioural traits consisted of an eight-item questionnaire rated on a three-point scale. Le Scolan *et al.* (1997) found significant correlations between the observed behaviours and temperament ratings. For example, the results from the arena test were positively correlated with gregariousness in familiar surroundings and unfamiliar surroundings. It was concluded that the two assessment methods were therefore measuring the same behaviours.

Summary

Personality assessment through behaviour tests and observations can be used to identify individual differences in animal behaviour responses. Short-term consistency in behaviour responses has been demonstrated but further work is required to explore why it is more difficult to demonstrate long-term consistency in behaviour. It has also been shown that behaviour assessments can be used to some extent to analyse the personality structure of animals (e.g. Svartberg and Forkman, 2002). Such analyses are, however, often restricted by the type of behaviours measured and as a result are likely to focus on those components linked to anxiousness and excitability. Finally the use of behavioural measurements and tests can be both time consuming and demanding of resources, thus limiting the number of animals that can be assessed and limiting the statistical power of the resulting analyses. It is clear that behavioural assessment has an important role within animal personality research; its limitations, however, mean that it is perhaps best used in conjunction with trait rating methods.

2.2 Evidence for animal personality

The previous sections of this review have explored the approaches used to assess personality, with specific focus on animal personality. It has, however, been assumed that animal personality is real and that it is quantifiable, but what evidence is there to support this assumption? Gosling and Vazire (2002) and Gosling (2001) have reviewed the subject in terms of methodology and reliability of data and found that in general, there is mounting evidence to support the existence of animal personality and that it is possible to quantify it. Gosling and Vazire (2002) discuss the evidence with reference to three major criteria derived from Kenrick and Funder's (1988) review of human personality literature. These criteria are; "*1) assessments by independent observers must agree with one another; 2) these assessments must predict behaviours and real-world outcomes; and 3) observer ratings must be shown to reflect genuine attributes of the individuals rated and not merely the observers' implicit personality theories about how traits covary*" (Gosling & Vazire, 2002, p. 608). These three criteria must be met in order to establish the existence of personality traits and the validity of personality assessment methods. The animal personality literature will now be reviewed with respect to these criteria. This section of the review aims to demonstrate how animal personality studies have met these criteria and to discuss the current evidence for animal personality and our ability to measure it accurately.

2.2.1 Assessments by independent observers must agree with one another

In both rating and behavioural assessments of animal personality, it is important to ascertain that the recordings are accurate. Factors that may affect accuracy can be organised into four categories; "*1) good judge, the possibility that some individuals might be better judges of personality than others; 2) good target, the possibility that some*

individuals might be more easily judged than others; 3) good trait (or behaviour), the possibility that some traits (and therefore some behaviours) might be easier to judge than others; and 4) good information, the possibility that more or certain kinds of information might make accurate judgements more likely" (Funder, 1995, p. 656). When considering Gosling and Vazire's (2002) Criterion One it is important that these factors be considered as they are all likely to influence observer (rater) agreement.

Inter-rater agreement is usually explored using multiple raters to assess the same group of animals. Correlation coefficients, for example, are then calculated between rater pairs to quantify the level of agreement between raters. Human rating studies have identified significant correlation coefficients of between 0.34 and 0.67 (Kenrick & Funder, 1988) and more recently of 0.24 and 0.54 (Funder *et al.*, 1995) with some variation in agreement levels between rater-type and acquaintance (Funder & Dobroth, 1987; Funder *et al.*, 1995; Gosling *et al.*, 1998). These values indicate that agreement between raters can account for up to 45% of variance. As it is generally accepted that human personality exists, these coefficients provide benchmark values for animal personality data to meet. Gosling (2001) summarised the correlation coefficients (and similar statistics) of inter-rater agreement of 21 animal personality studies. This provided an overall estimate of inter-rater agreement with a weighted grand mean coefficient of 0.52, which is comparable to that demonstrated in the human personality research. Although this generalisation of inter-rater agreement hides some of the variability in the animal personality data, it does demonstrate that inter-rater agreement in animal personality research is at least comparable to that of humans.

Inter-rater agreement is often assessed in association with trait reliability (good trait). Raters' scores for a given trait across the sample group are tested for association usually by using simple correlation analyses. This process helps to test the assessment tool whilst also

exploring the reliability of raters. Many authors, including Stevenson-Hinde and Zunz (1978), Stevenson-Hinde *et al.* (1980), Capitanio (1999), Caine *et al.* (1983), Feaver *et al.* (1986) and Gosling (1998) have used this approach and have demonstrated good levels of trait and rater reliability. The advantage of this method is that it allows for the refinement of trait lists by identifying 'good traits'. These can then be used to provide a reliable assessment of personality for the given species. Those traits that do not reach sufficient agreement are then removed from further analysis (e.g. FA or PCA) in order to improve the overall rigour of the data sample (Feaver *et al.*, 1986).

Alternatively, inter-rater reliability can be explored in association with the assessment of 'good targets'. In this case, raters' scores for each individual are considered and tested for association, rather than focussing on each individual trait. This approach allows for the identification of those individuals that raters find difficult to assess. This is a less common approach to rater reliability, but has been facilitated by Anderson *et al.* (1999) for assessing rater reliability in horses. Several parameters may affect the ease of rating a specific individual. These may include the level and type of acquaintance between raters and the individual being rated (Funder *et al.*, 1995; Gosling *et al.*, 1998; Gosling & Vazire, 2002), as well as the developmental stage of the animals being studied (Gosling & Vazire, 2002). Such parameters are an important aspect of animal personality and have yet to be explored in detail.

Finally inter-rater agreement is affected by the style of information recorded. With respect to rating data, this can be greatly affected by the type of rating scales applied. The more simple the rating system, the more likely raters are to find agreement. For example, Caine *et al.* (1983) used a three-point rating scale to assess the personality of ten pig-tailed macaques. A 93.3% level of agreement (between at least two raters) was demonstrated.

Although, not discrediting these results, the probability of raters agreeing on one score out of a possible three, is higher than that of raters agreeing for example, on a score out of seven. Thus by limiting choice, rater-agreement can be increased, but at the same time some detail as to the personality of an individual is lost. There is therefore a trade off between rater-agreement and the detail of the information gathered.

Morris *et al*. (2002a) explored rater reliability of nine raters who assessed the personality of ten horses. Each rater regularly handled the horses and had done so for an average of 2.8 years. The NEO-PI-FFI was used as the assessment tool and Kendall's coefficient of concordance (W) was used to explore rater agreement between all nine raters for each of the five components. Kendall's coefficients were significant ($P<0.02$) and ranged from 0.26 (*Openness*) to 0.62 (*Neuroticism*) indicating that concordance between raters was significant. Rater agreement was shown to be highest on the components *Neuroticism*, *Extraversion* and *Agreeableness*. *Openness* and *Conscientiousness*, however, were shown to be more difficult to rate. Rater agreement in this study focused on component scores as opposed to each horse's overall personality score. Although rater agreement was demonstrated to be quite high on a component level, it does not necessarily imply that the raters agreed on the overall scores for each horse.

In a study of 73 horses from five therapeutic riding schools, Anderson *et al*. (1999) assessed horse personality using a 20-item rating questionnaire and a five-point rating scale. Three regular handlers at each riding school rated the horses and agreement was assessed using Kendall's Tau-b rank correlation coefficients. Agreement between raters was relatively poor with agreement between two raters ranging from 29% to 50% of horses ($r > 0.52$, $P<0.01$). Variation was also identified as to which raters agreed the most. This

study focused on the agreement of overall personality of the horses as opposed to specific traits, but was unable to demonstrate high levels of inter-rater reliability.

In contrast, using a similar inter-rater agreement technique Martin (2005) was able to demonstrate much higher levels of reliability. Using an adapted version of the Stevenson-Hinde and Zunz (1978) trait list, Martin (2005) assessed the personality of 43 chimpanzees from five UK zoos. Each chimpanzee was rated by between three and five different raters, which included the regular handler and the author. Rater agreement was initially assessed using Kendall's Coefficient of concordance (*W*) to compare raters' scores for each individual. Raters were considered to have agreed if the *P* value for *W* was less than 0.01. Where *W* was not significant, data were entered into Spearman rank order correlations between rater pairs, to allow for identification of rater pairs that could not find agreement. Results indicated that of the 43 chimpanzees, 31 (72.1%) were rated reliably. Reliability of the ratings appeared to be related to how the chimpanzee had been reared, with mother-, group- reared individuals being more easily rated than those individuals that had been separated from their mothers as dependants and either reared alone or in a group with conspecifics.

Although it is important that trait reliability be quantified, the ability of raters to agree on the overall personality of the individuals being rated should also be taken into account. In other words good target should be assessed in addition to good trait and good judge. Furthermore, it should be noted that the rating method used could affect the rater-reliability.

2.2.2 Assessments must predict behaviours and real-world outcomes

In order for personality assessment to have a function and value, it must be able to predict behaviours and real-world outcomes (e.g. biological reactions or heritability). This is especially important in animal personality research, where personality assessment is thought to have potential applications in selection processes of animals for specific roles (e.g. Maejima *et al.*, in press).

In the reviews of animal personality literature by Gosling (2001) and Gosling and Vazire (2002), it is noted that few studies have clearly tested the personality measures used for the prediction of behaviour. In those studies that have explored this, the evidence for concurrent and predictive validity is strong (Gosling & Vazire, 2002), with personality-criteria correlation coefficients often exceeding the 0.30 level identified in the human literature (Mischel, 1968). The validity of animal personality assessment has been explored in two ways: correlations of personality ratings with conceptually related behaviours and correlations with real-world outcomes (Gosling & Vazire, 2002).

Personality-behaviour correlations explore the relationship between observer's trait ratings (or subsequent component scores) and coded behaviours. In a study of rhesus macaques, Capitanio (1999) explored personality behaviour correlations in contexts unrelated to those where the initial personality measurements had been made. Ratings of sociability were correlated with affiliative behaviours, and ratings of confidence were related to aggressive behaviours. Similarly in a study of cat personality, Feaver *et al.* (1986) demonstrated significant and strong correlations between personality ratings and behaviour. For example, ratings of aggressive were significantly correlated with aggressive behaviours (stare, hit and chase) and ratings of play were correlated with observations of play behaviour. Furthermore, Pederson *et al.* (2005) demonstrated that personality ratings can be used to

predict the behaviour of chimpanzees and found that agonistic context behaviours were positively correlated with the personality components *Dominance* and *Emotionality* and negatively correlated with *Agreeableness* and *Dependability*. *Extraversion* was positively correlated with affinitive context behaviours and negatively associated with public orientation behaviours. These were specific behaviours (aggressive display, explore, greet and watch) which were directed towards human observers, i.e. general public or zoo employees. The patterns of personality-behaviour correlations were consistent with the expected relationships. These examples provide evidence that observers are rating genuine attributes of individuals and not just anthropomorphic attributes (Gosling & Vazire, 2002).

Further evidence of links between personality ratings and biological mechanisms, have been identified in studies exploring the heritability of personality (Weiss *et al.*, 2000; Gauly *et al.*, 2001; Dingemanse *et al.*, 2002; Weiss *et al.*, 2002; van Oers *et al.*, 2003; Fairbanks *et al.*, 2004) and have been reviewed by Bouchard and Loehlin (2001), Plomin *et al.* (1994) and van Oers *et al.* (2005). Using bi-directional breeding studies, heritability has been demonstrated in the exploratory behaviour of great tits (*Parus major*) (Dingemanse *et al.*, 2002; Drent *et al.*, 2002) with individuals being described as fast or slow explorers in an open-field test. Furthermore, these differences in personality type have been tested for consistency and have been explored in relation to evolutionary theory (Marchetti & Drent, 2000; Dingemanse *et al.*, 2002; Dall, 2004; Both *et al.*, 2005; Carere *et al.*, 2005). In addition to the detailed studies on great tits, heritability of personality has also been demonstrated in dogs (Goddard & Beilharz, 1983; Strandberg *et al.*, 2005), farmed mink (Malmkvist & Hansen, 2002) and chimpanzees (Weiss *et al.*, 2000; Weiss *et al.*, 2002). These examples provide further evidence of the links between personality ratings and real-world outcomes.

In a study on horses Wolff *et al.* (1997) were able to test for genealogical effects by assessing horses that were paternal half siblings. The results identified that more offspring of stallion A tended to show 'excited' behaviours like passage and raised tail, whereas the offspring of stallion B showed more 'calm' behaviour such as exploration. This indicates that Wolff *et al.* (1997) were measuring real differences in behaviour and that these differences were controlled, to some extent, by genetics. Moreover, these results provide further evidence that animal personality research is capable of identifying individual differences and linking these to real-world outcomes, in this case, the inheritance of specific personality traits.

Personality data that are supported by theories of behavioural ecology, or by known natural history of a species, can provide further validity of animal personality assessment. For example, Gosling (1998) found that observer ratings of assertiveness in spotted hyenas were related to dominance status and sex. Dominant individuals were rated as more assertive, and females were rated as more assertive than males. This is consistent with the matriarchal dominance hierarchy found in hyena clans. Furthermore, no other traits were found to significantly correlate with dominance status or sex, thus demonstrating a link with a real-world outcome.

Similarly, Wielebnowski (1999) assessed 44 adult captive bred cheetahs and related personality scores to reproductive success. Observer ratings of *Tense-Fearful* could be used to differentiate between breeders and non-breeders. Non-breeding cheetahs were rated higher on the *Tense-Fearful* component than breeding cheetahs, but scores did not differ on any other trait for these two groups. This is consistent with previous observations of the adverse effects of anxiety on reproductive success (Boissy, 1995).

Strong relationships between personality and behaviour and real-world outcomes have been identified in a variety of different species. In contrast, Visser *et al.* (2003a) were unable to predict show jumping performance by using a variety of different behavioural tests on young horses trained for show jumping. Although some personality-behaviour links were identified, prediction of performance was not consistent and further refinement of the behaviour tests was recommended. It was noted that the tests used might not have assessed those qualities that were most predictive of show-jumping performance. This is one of the limitations of such behavioural tests, in that they are restricted to the characteristics of personality and temperament that they can measure. It seems that stronger personality-behaviour correlations have been identified in those studies that have used trait-rating techniques (e.g. Gosling, 1998; Wielebnowski, 1999; Pederson *et al.*, 2005; Maejima *et al.*, in press).

Animal personality studies have so far demonstrated logical relationships between personality and a variety of different biological functions and mechanisms as well as behaviour observations. These studies therefore, demonstrate that Criterion Two can be met by animal personality research and perhaps more so in those using trait rating methods of personality assessment.

2.2.3 Observer ratings must be shown to reflect genuine attributes of the individual

Criterion Three requires ratings to reflect attributes of the target animals and not the observers' own beliefs or theories. Many of the animal personality studies that have been published have demonstrated the existence of personality factors in the focal species (reviewed by; Gosling & John, 1999; Gosling, 2001; Gosling & Vazire, 2002). Several of these factors or dimensions have been shown to occur in many different species and have

also been repeatedly shown for the same species (see; Gosling & John, 1999; Gosling, 2001; Gosling & Vazire, 2002). In trait rating studies, however, it is possible that the researchers are not necessarily detecting personality structure, but are instead using their implicit knowledge of personality to mould the data accordingly. In contrast studies based on behavioural coding, tests and ethological observations are less prone to this. Sinn *et al.* (2001) studied 73 baby octopuses and identified four factors which were labelled *Active Engagement* (behaviours such as touch stimulus, crawl, colour change, and jet or swim), *Arousal/Readiness* (behaviours such as head move, respiratory change, and papillary change), *Aggression* (behaviours such as grab brush, pull brush and posture change) and *Avoidance/Disinterest* (behaviours such as papillae change and shrink). With the exception of *Arousal/Readiness*, these results mirrored those found by earlier trait based research on octopuses (Mather & Anderson, 1993). Such behaviour-based factors cannot be attributed to the observers applying their own theories to the data (Gosling and Vazire, 2002). Additionally, in cross study comparisons, behaviour rating studies often correlate with the findings of trait rating studies suggesting that both methods are reliable and are assessing the same phenomenon (Gosling and Vazire, 2002).

Those studies, previously discussed with respect to Criterion Two that demonstrate personality-behaviour correlations are also of relevance here. By demonstrating logical links with behaviour it can be implied that observers are providing genuine assessments of the individuals being studied. For example, the study by Capitanio (1999) on the personality of rhesus macaques, not only found personality-behaviour correlations in those contexts where the initial personality assessments had been made, but also across different social contexts. This indicated that the personality scores assigned to each individual were consistent across different social contexts and were genuine assessments of each individual. These and other studies that have demonstrated links between personality and

behaviour (Feaver *et al.*, 1986; Pederson *et al.*, 2005) offer further evidence that observers are providing genuine assessments of personality. It can therefore be concluded that animal personality research has met Gosling and Vazires's (2002) third criterion.

Summary

Gosling and Vazire (2002) concluded that animal personality studies are able to satisfy all three of Kenrick and Funder's (1988) criteria for reliable and valid personality assessment and therefore demonstrate not only the existence of personality in animals, but also our ability to measure it accurately. In addition, they demonstrate that animal personality studies are comparable in reliability and standard to that of human studies. Personality ratings of animals have shown strong levels of inter-observer reliability and in some cases these assessments show validity in terms of predicting behaviours. Such assessments do not merely reflect the implicit theories of observers projected onto animals (Gosling and Vazire, 2002). It is clear, however, that further research, particularly on the factors that affect rater reliability, is still required in order to improve and standardise assessment methods and reliability.

2.3 Approaches to horse personality assessment

The study of individual differences and personality in horses is an increasingly common topic of research. Recent studies have explored the assessment of individual differences through the use of behaviour and temperament tests (e.g. Le Scolan *et al.*, 1997; Wolff *et al.*, 1997; Visser *et al.*, 2001; Seaman *et al.*, 2002; Visser *et al.*, 2002; 2003a; 2003b) and through the use of ratings provided by handlers (Anderson *et al.*, 1999; Creighton, 2003; Momozawa *et al.*, 2003; 2005). Such studies have demonstrated that reliable assessments of individual differences are achievable and the subject has previously been reviewed by Mills (1998) who concluded that assessment methods had potential applications within the equine industry.

Behavioural assessment of horse personality has utilised behavioural tests such as the novel object, arena, handling and learning tests (e.g. Le Scolan *et al.*, 1997; Visser *et al.*, 2001; Seaman *et al.*, 2002; Visser *et al.*, 2002; Hausberger *et al.*, 2004; McCall *et al.*, 2006). These tests have been able to demonstrate individual differences in behaviour but have only demonstrated limited consistency over time (Visser *et al.*, 2001; Seaman *et al.*, 2002). Behavioural assessments have also been carried out alongside biological indicators (e.g. heart rate) and ratings made by handlers. For example, Le Scolan *et al.* (1997) assessed 72 horses using four behavioural tests. These were arena and novel object tests, instrumental and spatial learning tests and a memory test. In addition to the behaviour tests, the horses were assessed using a simple eight-item rating questionnaire with a three-point rating scale. Items included 'fearful when ridden' and 'nervous when handled'. Several rating-behaviour correlations were identified. These included; a correlation between the rating of gregarious and the reactivity in the arena test and reactivity in the novel object test with the rating of nervousness when ridden. The authors concluded that the

behavioural tests were good predictors of the horses' overall temperament and personality. Similarly Visser *et al.* (2002) demonstrated links between reactions in novel object and handling tests with heart rate (HR) and heart rate variability. Their results identified that these tests brought about HR responses in the horses and may be utilised as indicators of a horse's temperament. McCann *et al.* (1988) were also able to demonstrate links between personality measurements, rated during a chute test, and HR measured at different points throughout the test.

As previously discussed, the efficacy of trait-rating assessments has already been demonstrated in a variety of species (see Section 2.1.3) and is the main method of assessment in humans for both peer and self-assessment (Pervin & John, 1997). Trait rating methods have also been applied in various forms for equine personality assessment. For example, Morris *et al.* (2002a) employed an in-depth human personality rating system, whilst French (1993), Anderson *et al.* (1999) and Momozawa *et al.* (2003; 2005) used behaviourally defined adjectives to create rating questionnaires. In contrast Le Scolan *et al.* (1997) scored situation specific behaviours (e.g. fearful when ridden) in conjunction with behavioural assessments. The reliability of these assessments, however, has been varied. For example, Anderson *et al.* (1999) were unable to demonstrate high levels of inter-rater reliability, and postulated that the horses may have been behaving differently for each of the handlers. The relationship between handlers and horses was not specified, thus it is possible that some of the handlers had not known some of the horses for very long, and were therefore unable to provide accurate assessments of the horses. In contrast, Morris *et al.* (2002a) demonstrated high levels of inter-rater reliability when they applied the NEO-PI-FFI, a human personality rating tool, to the assessment of ten horses by nine raters. Each of the nine raters had worked with the group of horses for an average of 2.8 years.

This long acquaintance with each horse is likely to have increased the accuracy of the judgements.

The use of trait rating assessment must also be able to demonstrate that ratings represent genuine assessments of personality and that they are related to real world outcomes and behaviour (Gosling & Vazire, 2002). Momozawa *et al.* (2003) demonstrated links between personality scores on an eight-item questionnaire with the results of a reactivity test. For example, horses that had been scored as highly anxious tended to show greater HR increases and defecated more often during the exposure to the reactivity test, than did those rated as low on *Anxiousness*. Le Scolan *et al.* (1997) were also able to demonstrate links between rated scores of temperament with behaviours measured during four behaviour tests. For example, reactivity in the arena test was found to be associated with the score for gregariousness.

Momozawa *et al.* (2005) have further demonstrated the validity of questionnaire-based ratings for horse personality assessment. They demonstrated consistency in factor structure over two years, using two sample groups of horses (n = 69 and 70) and a 20-item questionnaire rated on a nine point scale. Three handlers rated each horse and their average scores were entered into a PCA, which extracted five factors in both years, explaining 71.4% and 75.5% of the variability. Four of these factors were shown to be stable across the two years. This study did not, however, test for inter-rater reliability or links with behaviour or real-world outcomes.

To demonstrate the reliability of horse personality assessment it is important that research meets the three criteria defined by Gosling and Vazire (2002). To date horse research has met some of these criteria, but few studies have clearly met all three. Morris *et al.* (2002a)

demonstrated significant levels of inter-rater agreement (Criterion One), but did not appear to meet Criteria Two or Three. In contrast Visser *et al.* (2003a) explored the relationship between personality and behaviour by demonstrating the potential for one to predict the other (Criterion Two). Behavioural expression of personality was measured in young horses and was compared to jumping performance. The study provided some indication of a link but emphasised the need for different behavioural tests to assess relevant personality traits that might determine an individual's success as a show jumper. Finally, Momozawa *et al.* (2003) and Le Scolan *et al.* (1997) demonstrated that raters were able to rate genuine attributes of the individuals by demonstrating that personality scores were significantly correlated with behaviours measured during behaviour tests (Criterion Three). Furthermore, Momozawa *et al.* (2005) provided evidence of the consistency of horse personality ratings by demonstrating stability of factor structures over two years.

To date, no horse personality research has employed the Stevenson-Hinde *et al.* (1980) trait list for the assessment of personality in horses. As previously discussed, the Stevenson-Hinde *et al.* (1980) trait list has been successfully adapted to a wide variety of animal species including cats (Feaver *et al.*, 1986), gorillas (Gold & Maple, 1994), spotted hyenas (Gosling, 1998), cheetahs (Wielebnowski, 1999), chimpanzees (Martin, 2005) and pig-tailed macaques (Caine *et al.*, 1983). This approach to personality assessment is flexible due to the use of behaviourally defined adjectives that can be easily applied to other species. In addition the assessment method has been used in conjunction with PCA to identify underlying personality components and explore personality structure. Adaptation of this method for the assessment of horse personality should provide reliable results and provide an indication of the factor structure of horse personality. Subsequent factor structures could be compared to other species allowing for a more in-depth understanding of personality evolution and development. It therefore seems possible that this assessment

tool has potential to be applied for the assessment of personality in horses. It is, however, important that the development of a novel assessment scale for horse personality follows the appropriate development stages so as to ensure both reliability and validity of the final scale.

2.3.1 Development of horse personality assessment scales

The previous sections have had a significant focus on the general approaches to personality assessment and explored the use of some of the key theories/concepts. The process of developing such assessment methods, however, has yet to be discussed. It is important to understand such processes as they provide the foundations of personality theory. The development of personality assessment methods should, therefore, build on previous theory and knowledge as well as empirical research (Ramsay & Reynolds, 2000). Furthermore, as with all behavioural and psychological research, reliability and validity must be considered during the developmental stages (Martin and Bateson, 1993; Dytham, 2003; Pervin, *et al.*, 2005)

The key stages involved with the development of psychometric scales and personality assessment methods, were discussed in detail by Ramsay and Reynolds (2000) who suggested the following eight steps: 1) Review the literature; 2) Define the construct; 3) Test planning and layout; 4) Designing the test; 5) Item try-out; 6) Item analysis; 7) Building a scale; and 8) Standardising the test.

Within the animal personality literature, there are a range of different approaches used, it is important, therefore, to assess whether such methods have been developed in an appropriate manner. These are now discussed further, in relation to Ramsay and Reynolds' (2000) eight stages of development.

Stage one: Review the literature

Ramsay and Reynolds (2000) identified that, as with any type of scientific research, the reviewing and critical analysis of research literature should be the first stage in the development of a psychometric test, such as those used for personality assessment. This allows the test designer to identify how other researchers have measured the construct under assessment (e.g. personality). Such analysis helps to build an understanding of what is already known, furthermore, approaches that have been used in the past can be adopted and further improved, thus adding additional evidence of their reliability and validity (Ramsay and Reynolds, 2000).

Such an approach was clearly demonstrated by Morris *et al.* (2002a; 2002b) who provided an extensive literature review which evaluated not only the concept of animal personality research, but also evaluated previously adopted methods. In particular, Morris *et al.* (2002a) acknowledged previous criticisms made by Gosling and Bonnenburg, (1998) as to the problems of small sample sizes and the lack of a standard taxonomy in terms of animal personality assessment. Furthermore, upon the analysis of previous horse personality studies, Morris *et al.* (2002a) identified flaws in earlier assessment methods. For example, Mills (1998) was noted for not providing respondents with definitions of selected traits/assessment items, therefore allowing respondents to provide their own interpretations. This, in turn, was thought to have resulted in non-significant inter-rater agreement. In contrast, Morris *et al.* (2002a) acknowledged the success of using a peer rating system to assess human personality (e.g. Costa & McCrae, 1992) and therefore provided justification for exploring such a phenomenon in animal personality assessment. Thus, the review by Morris *et al.* (2002a) clearly demonstrated the thought processes involved with the development of their assessment method.

Unfortunately, such detailed reviews of the literature are not always provided within published works. Thus a clear understanding of the thought processes involved with the development of assessment methods may not always be provided. For example, Momozawa *et al.* (2003), Le Scolan *et al.* (1997) and Anderson *et al.* (1999) all provided limited reviews of the research literature. Anderson *et al.* (1999) in their assessment of riding school horses did, however, utilise a modified form of a previous rating scale used for the assessment of equines (French, 1993). They did not, however, provide clear justification of either the use of this scale or its adaptation by the addition of further adjectives. It would appear that studies often utilise previous rating methods, but do not always clearly acknowledge the thought processes involved in justifying their use and/or modification. Future assessment scales should, therefore, be developed in light of research literature and take into account the criticisms of previous methods.

Stage two: Define the construct

In any area of research it is highly important to clearly define what is to be studied. This helps to prevent confusion with similar or related terms. With respect to individual differences, the terms personality and temperament are often used in the literature to describe this phenomenon (see Section 2.1.4). These terms are not, however, always defined clearly in the published literature, resulting in the potential for confusion over what is actually being assessed. Furthermore, the similarity in their definitions adds further confusion as to whether or not they are in fact measuring the same construct.

With respect to the horse personality literature, both temperament (French, 1993; Le Scolan, *et al.*, 1997; Visser *et al.*, 2001; Momozawa *et al.*, 2003; 2005;) and personality (Mills,1998; Morris *et al.*, 2002a, 2002b) are readily used to describe individual differences with the term temperament used in connection with both behavioural and trait

based assessment methods. For example, Momozawa *et al.* (2003, 2005), Anderson *et al.* (1999) and French (1993) assessed temperament using questionnaire/trait-based methods.

Few of the studies on individual differences in horses have clearly defined their selected terminology. For example, Momozawa *et al.* (2003; 2005) developed and tested a questionnaire based assessment method for measuring the temperament of horses and despite having discussed the practicalities and potential applications they did not provide a clear definition of what temperament was. Similarly, Morris *et al.* (2002a), in their study using the NEO-PI-FFI for the assessment of personality in horses, failed to provide a clear definition of personality. They did, however, attempt to differentiate between personality and temperament, stating that the latter is that which is apparent from birth. Without a distinct definition of personality, this seemed to be a very one sided comparison and of limited benefit.

Anderson *et al.* (1999) confuse things further by stating that in their temperament survey, 20 personality traits were used in order to quantify temperament. Furthermore, Anderson *et al.* (1999) did not define either temperament or personality and later also referred, without providing justification, to emotionality as being part of the same construct. This interchange of terms without justification or explanation makes the interpretation and comparison of such studies far more complex.

In contrast, some studies have provided clear and justified construct definitions, providing a range of definitions and discussion of their validity. For example, Le Scolan *et al.* (1997) provided a definition of temperament early on in their study and later go on to identify how this relates to methods of assessment. Similarly, Visser *et al.* (2001) provided a clear definition of temperament and highlighted the key concepts of this term (e.g. consistency

across situations). Although both of these studies clearly identified temperament, neither provided a contrast to personality, nor discussed the similarity and close relationship of these terms. More recently McGrogan *et al.* (in press) provided a more evolutionary based definition of personality which identified the stability of personality over the lifespan of an individual. They also highlighted the view of both Sih *et al.* (2004b) and Wolf *et al.* (2007) that individual differences reflect adaptively relevant individual differences in trade-offs among trait values. Such a definition, although enlightening, does not, however, provide a clear definition of the construct being assessed.

In summary, within the horse personality literature clear definitions of the construct being assessed are not always provided, thus making valid contrasts between studies more difficult. Future research should, therefore ensure that constructs are clearly defined and justified, thus ensuring that the resulting methods of assessment are appropriate.

Stage three: Test planning and layout

The third stage recommended by Ramsay and Reynolds (2000) identified the importance of the selection of representative samples in terms of the behaviours selected for assessment. In other words it explored what Anastasi (1988) termed the *behaviour domain*. This included the range of behaviours that a test claims to measure, and describes the total range of behaviours as being equivalent to a population and those behaviours that are selected for measurement, as being the behaviour sample. Thus a behaviour sample may be extensive and utilise a wide range of items, or may be small and therefore more selective as to which behaviours are being measured. Either way, the behaviour sample should be reflective of the behaviour domain (Ramsay & Reynolds, 2000). Therefore, to meet such a goal, a personality test, for example, should assess all of the important aspects of the characteristic or characteristics of interest (Ramsay & Reynolds, 2000)

When this is taken into account with respect to the horse personality literature, several assessment measures appear to be limited in the characteristics they are measuring. For example, Le Scolan *et al.* (1997) only used seven questions to assess horse behaviour/temperament. Similarly, Momozawa *et al.* (2003) used eight traits in their horse temperament questionnaire, but later extended their list to 20 (Momozawa *et al.* 2005). Such assessments are very specific and may not be able to tell us much about the general personality structure of horses.

In contrast Morris *et al.* (2002a) provided an extensive assessment tool that used 59 questions from a human personality rating tool (NEO-PI-FFI) and covered five personality factors/dimensions. The relevance, however, of this assessment method has already been questioned earlier in this review (Section 2.1.3) due to its distinctly humanised questions. The behaviour sample used by Morris *et al.* (2002a) could be considered as an inappropriate reflection of the population in terms of horse personality.

McGrogan *et al.* (in press) adopted an alternative approach to assessing horse personality by adapting a method used by Ley *et al.* (in press) for the personality assessment of dogs. The personality adjectives (or traits) were selected by a total of 30 regular handlers of the horses and were, therefore, thought to reflect the broader dimensions of horse personality as well as be easily interpreted by horse handlers. Such an approach led to an extensive list of 36 traits that provided a broad behaviour sample.

With the exception of McGrogan *et al.* (in press) horse personality studies are generally in a distinct contrast with the assessment tools utilised by other animal personality researchers (see Section 2.3), many of whom have adopted the 30 traits provided by Stevenson-Hinde *et al.* (1980). These traits provide a practical and yet extensive assessment of animal

personality and provide an appropriate behaviour sample for a wide range of species as discussed in Section 2.1.3 (*Trait rating approach in animals*).

Stage four: Designing the test

The design layout and structure of the final test is important in terms of allowing for ease of use and interpretation, as well as avoiding bias or misleading the rater (Ramsay & Reynolds, 2000). The inclusion of clear instructions and the use of an appropriate item type also allow for more accurate results. With reference to personality assessment, item types can be classified into four categories, dichotomous, rating-scale, midpoint and forced choice, furthermore, each of these types may come in various forms (Ramsay and Reynolds, 2000).

A dichotomous item limits the rater to only two options, for example, yes or no, or true or false. Although well used within human personality research (See Ramsay and Reynolds, 2000) such an approach does not appear to have been utilised within animal personality. This may be due to limitations in the amount of information that can be extracted.

In contrast, a typical rating-scale item gives the respondent greater flexibility due to the responses falling upon a rough continuum (Ramsay and Reynolds, 2000). Rating scales (or Likert scales, Coolican, 2004) often request that the respondent rates an individual using a fixed scale often between one and five, or one and seven (referred to as a Likert scale; Coolican, 2004). Where one represents minimal or negligible expression of the trait, and the highest value represents full expression of the trait in question. Alternatively each response may be individually labelled, for example, rarely, sometimes, often and very often. Items rated with odd numbers have mid points, with the central point often used to represent a neutral or average score.

A large volume of animal personality studies have utilised such rating scales, the number of options provided however, varies a lot between studies. For example, Stevenson-Hinde and Zunz (1978), Stevenson-Hinde *et al.* (1980), King and Figueredo (1997) and McGrogan *et al.* (in press) all utilised a seven-point scale. Whereas both Caine *et al.* (1983) and Le Scolan *et al.* (1997) employed a three-point scale and McCann *et al.* (1988) used a four point scale. Few studies have demonstrated any major consideration as to why such scales have been selected, although Momozawa *et al.* (2003, 2005) did show the development process of their assessment method, using a five-point rating scale in 2003 and later expanded this to a nine-point scale (Momozawa *et al.*, 2005), but did not provide clear justification for this modification.

In contrast to the limitation of the traditional rating scale, both French (1999) and Morris *et al.* (2002a) utilised labelled analogue scales which allowed the respondent greater flexibility in their ratings, thus resulting in a more continuous data set. This method can, however be rather laborious with respect to transforming the marks into real numerical values and has not been widely used in animal personality assessment.

Finally, forced choice items usually present paired sentences and the respondent is forced to select the sentence that best describes the individual being rated (Ramsay & Reynolds, 2000). This item type does not appear to have been utilised in the assessment of animal personality with researchers tending to select the fixed rating scale as the preferred item type.

In addition to the selection of appropriate item-types, Ramsay and Reynolds (2000) suggested steps be taken prior to implementation of the final test. Firstly that the test should be reviewed both by colleagues and the authors themselves for both clarity and

sense. It may also be beneficial to have relevant peers review the terminology used within the test. With reference to animal personality studies, such consultation should ideally be sought with those individuals that have a large amount of experience with the target species, so as to allow for both representative and relevant phrases. Such consultation was acknowledged by Stevenson-Hinde and Zunz (1978), Stevenson-Hinde *et al.* (1980) and McGrogan *et al.* (in press) in the development of their personality rating methods. Furthermore, such studies have incorporated clear behavioural/dictionary definitions of the personality terms used in order to aid consistency of interpretation. Moreover, Ramsay and Reynolds (2000) also suggested that authors avoid the use of complicated language and the use of complex grammatical constructions.

In addition, Ramsay and Reynolds (2000) identified that the overall organisation of the questionnaire/test is of great importance in that it can influence the test's effectiveness in measuring the desired characteristics. For instance, the order in which traits are included should allow for the more positive items to be included early on and that any disturbing or negative items should be re-phrased or replaced where possible (Ramsay & Reynolds, 2000). Such considerations do not appear to be acknowledged within the animal personality literature with traits seemingly ordered in relation to their appropriate factors (Morris *et al.*, 2002) or in seemingly random orders (for example, Le Scolan *et al.*, 1997; Anderson *et al.*, 1999; Momozawa *et al.*, 2003, 2005; Pederson, *et al.*, 2005; McGrogan, *et al.*, in press). It is not clear, however, what affect this may have had on the results of such studies but should be taken into account during the development of new assessment methods. Ramsay and Reynolds (2000), however, acknowledge that the effects of test layout have received limited research attention.

Stage five: Item tryout

Upon completion of the previous four stages, Ramsay and Reynolds (2000) identified that the test developer should reach a stage where the assessment can be tested on a representative sample of the population. Such a try-out sample should aim to match the characteristics of the final target population. The aim of the try-out phase is to evaluate the validity of the test as well as its overall functionality.

Such try-out or pilot studies are rarely recorded within the published animal (or horse) personality literature. Furthermore, the impression is sometimes given that published research is actually part of this particular development stage. Some studies do, however, relate back to earlier work by the same authors to identify how the assessment tool was previously developed. For example, Stevenson-Hinde *et al.* (1980) and Momozawa *et al.* (2005) both refer back to the results of their earlier studies (Stevenson-Hinde & Zunz, 1978; Momozawa *et al.*, 2003, respectively) in order to identify and justify the developments and improvements made to their assessment tools in-between publications (see Stage Six for further details). Analysis of pilot data and the refinement of the assessment tool are discussed further under Stages Six and Seven.

Stage Six: Item analysis

Stage six explores the ability of the scale to discriminate between individuals and to provide as much information as possible about the differences between individuals on the characteristic being measured (Ramsay & Reynolds, 2000). Such that individuals with high levels of a characteristic should score highly and those individuals with low levels of a characteristic should receive low scores. Furthermore, the distribution of scores should be relatively spread out across the range of possible scores, as opposed to clumping at one or

two points (Kline, 1993b). A test's ability to demonstrate such distinction between individuals is often referred to as its discriminating power (Kline, 1993b).

The difficulty and the attractiveness of an item are thought to affect its discriminating power (Ramsay & Reynolds, 2000). If an item is too difficult to rate then even those individuals who are fairly high on that item are likely to be scored incorrectly. Thus reducing the efficiency of the test in discriminating between high and low individuals on the characteristic being measured. In human personality research an item-difficulty (or attractiveness) index can be calculated which indicates the proportion of respondents who gave the desired (or keyed) response for a particular characteristic, such that their response matches their actual level on that characteristic (Ramsay & Reynolds, 2000). Such measures are difficult to achieve with animal personality research partly because raters are assessing the perceived personality of an individual and, as few alternative measures of personality are available, it is difficult to ascertain if a score has been accurate.

Within animal personality research, item difficulty appears to be assessed through inter-rater reliabilities. For example, those items that do not receive significant agreement between multiple raters are deemed difficult to understand and therefore unreliable. It is assumed that in such cases, raters are interpreting traits differently. One method of increasing agreement between raters and increased comprehension of traits is to involve potential respondents in the development of item or trait definitions. This approach was recently utilised by McGrogan *et al.* (in press) who involved 30 experienced horse handlers in the development of their horse personality questionnaire. Traits were then defined using adaptations of the dictionary definitions. Those traits found to have low reliabilities between raters were then removed from further analysis. This process has also been used by Feaver *et al.* (1986) and Stevenson-Hinde and Zunz (1978) in order to improve the

robustness of their scales. In contrast Morris *et al.* (2002a) used human based rating terms, some of which were difficult to relate to horse behaviour. This may have made the interpretation of these terms more difficult. Despite this the authors reported acceptable levels of reliability. These were noticeably lower, however, on the components which utilised those traits that were more difficult to relate to horses, for example those associated with *openness*.

The evaluation of items for their ease of use is an important stage of development and is strongly associated with the eighth development stage described by Ramsay and Reynolds (2000) which looks to standardise the test (see later). The selection of appropriate traits that can be easily understood and related to horse behaviour, is therefore, an important stage in the development of a new horse personality assessment method.

Stage seven: Building a scale

The process of item analysis allows for the identification or modification of unreliable traits/terms such that only those terms judged to be acceptable are entered into further analysis. The final scale/assessment tool then needs to be built such that the final traits are organised in an appropriate manner, for example by personality factor or characteristic (Ramsay & Reynolds, 2000). As discussed earlier (Section 2.1.3) factor analysis and PCA are commonly used in both human and animal personality assessment to group closely related items together into dimensions or factors. Such an approach was justified by Stevenson-Hinde and Zunz (1978) for use with animal personality due to its independence from previous theories as to how the traits should organise themselves. This was in contrast to earlier studies which had assumed that animal personality structure would match that of humans (See Sections 2.1.2 and 2.1.3). Kline (1993b), however, advised careful use of such analyses stating that, even after assessing the attractiveness and

discriminating powers of individual items, it is still possible to obtain a test that measures the wrong characteristic reliably and discriminably. Further discussion as to the use of factor analysis is discussed under stage eight.

Stage eight: Standardising the test

The aim of Ramsay and Reynolds' (2000) final stage is the standardising of the test and the demonstration of its reliability and validity. A test demonstrates reliability when it shows consistency across time, conditions, scorers, items or test forms (Ramsay and Reynolds, 2000), such that observations can be replicated (Pervin *et al.*, 2005). In essence, reliability is freedom from error (Ramsay and Reynolds, 2000). Test validity, however, is concerned with whether or not the test is measuring what it is meant to (Ramsay and Reynolds, 2000) such that it explores the extent to which observations actually reflect the phenomena under investigation (Pervin *et al.*, 2005).

Pervin *et al.* (2005) identified that test reliability can come in two forms. Firstly, internal consistency which assesses the extent to which the different items on the test correlate with each other and whether this is what would be expected if each item was a reflection of a common psychological construct. The second form looks at test-retest reliability, this is a measure of consistency across time and assesses, for example, how well an individual's scores correlate if they have been assessed at different points in time.

With respect to the animal personality literature, internal consistency of the resulting scales or factor models is not commonly explored or reported. Stevenson-Hinde and Zunz (1978) and Stevenson-Hinde *et al.* (1980) did, however, explore internal consistency by assessing stability of their model structure over four successive years. Their analysis identified relative stability across years, in both structure and individual scores. With the addition of

new adjectives, however, in the third and fourth years a third component was identified, but the overall structure of the model was in keeping with the previous two years.

Internal consistency was explored in greater depth by Momozawa *et al.* (2005) in their assessment of horse personality of 139 thoroughbreds across two years. They used a 20-item questionnaire (each item rated from one to nine) to survey 69 horses in 2002 and 70 horses in 2003. Data were analysed for each year using principal factor analysis with a varimax rotation and initially extracted five factors. Following the use of Cronbach's alpha (a measure of internal consistency) they identified that 15 of the 20 items reliably loaded onto their appropriate factors, but that only three factors demonstrated sufficient internal consistency. Similarly, in their assessment of horse personality using the NEO-PI-FFI, Morris *et al.* (2002a) explored internal consistency by measuring the degree to which individual items contributed to the overall factor scores. Morris *et al.* (2002a) demonstrated that *Neuroticism* and *Extraversion* showed the greatest stability out of the five factors assessed. With the exception of these studies and the recent investigation by McGrogan *et al.* (in press) horse personality studies have not explicitly assessed the internal consistency of the resulting personality components.

Within the animal personality literature test-retest reliability on an individual's score is more commonly assessed using multiple raters and then looking for agreement of scores (as was discussed in Section 2.2.1). For example, Stevenson-Hinde and Zunz (1978) and Stevenson-Hinde *et al.*, (1980) used a minimum of three raters to assess each of the rhesus macaques in their study and subsequently used Pearson correlations to assess the agreement between rater pairs on each adjective/trait. Agreement was said to have been acceptable when $P<0.05$. Traits which did not reach this level of significance were subsequently removed from further analysis. This method was later adopted by Feaver *et*

al. (1986), Capitanio (1999), Caine *et al.* (1983) and Gosling (1998). The removal of unreliable traits is thought to improve the overall reliability and vigour of the test being constructed (Feaver *et al.*, 1986).

Such a process has been demonstrated in some horse personality studies, for example, McGrogan *et al.* (in press) employed intra-class correlations to assess inter-rater reliabilities and internal consistency. This process identified two adjectives (stupid and wise) that were found to have poor reliability between raters and were removed from further analyses. In contrast, Anderson *et al.* (1999) explored the reliability of raters to assess horses on 20 paired personality adjectives but failed to show good inter-rater reliability. The authors did not, however, follow this with the removal of the most unreliable traits, and therefore did not report any attempts to improve their assessment tool. Nor did they explore which items had shown the greatest unreliability. These data were not entered into factor analysis or PCA, thus internal consistency was not measured.

In contrast to McGrogan *et al.* (in press), Morris *et al.* (2002a, 2002b) and Anderson *et al.* (1999), the majority of horse personality literature does not report the use of multiple raters in order to assess the reliability of their assessment tools. Thus the reliability of their selected adjectives/terms to produce repeatable results has not been demonstrated.

In addition to reliability, the validity of a test must also be demonstrated (Ramsay and Reynolds, 2000). For a test to have construct validity evidence is required that shows the test is indicative of the psychological construct being measured (Pervin *et al.,* 2005). This can be demonstrated in a number of ways for example, convergent and divergent validity (Anastasi, 1988). Convergent validity seeks to demonstrate that a test behaves similarly to other tests that purport to measure the same or similar constructs. Conversely, a test shows

divergent validity when it behaves differently to tests of dissimilar constructs (Anastasi, 1988). It is therefore important that a clear and relevant construct definition be defined in the earlier development stages (i.e. at Stage two) so that the ability of the test to measure the defined construct can be clearly assessed.

Both Pervin *et al.* (2005) and Ramsay and Reynolds (2000) state that psychologists can demonstrate construct validity by showing that the test relates systematically to some external criterion or non-test real-world outcomes. Such external tests should, however, be selected using theoretical considerations (Pervin *et al.*, 2005). This is similar in context to the third criterion described by Gosling and Vazire (2002) (that assessments must predict behaviours and real-world outcomes).

As discussed in Section 2.2.2 a number of animal personality studies have successfully identified links between personality ratings and behaviour or real-world outcomes (e.g. Capitanio, 1999; Weiss *et al.*, 2002; Dingemanse, *et al.*, 2002; Drent *et al.*, 2002; Pederson *et al.*, 2005). Similar links have also been identified in horse personality studies, for example Wolff *et al.* (1997) identified links between personality and the genealogy of the horses that they assessed, such that individuals with the same paternity tended to have more similar personalities. Momozawa *et al.* (2003) also compared their personality ratings to behaviour measured during a balloon reactivity test and found significant correlations between personality scores and change in heart rate measured during the test.

In contrast, Visser *et al.* (2003a) were unsuccessful in using personality (assessed using behaviour tests) to predict jumping ability, and identified that their tests may not have been assessing the qualities that were required for the prediction of show jumping performance.

Thus their definition of the construct being assessed may not have been sufficiently clear or accurate and may have resulted in reduced validity of their tests.

In summary, within the horse personality literature, there is limited evidence of researchers fully demonstrating the validity of their assessment methods (e.g. Anderson, 1999; Morris *et al.*, 2002a; McGrogan, *et al.*, in press). Therefore, future horse personality assessment methods need to demonstrate strong validity.

Summary

Upon reviewing the animal and horse personality literature with respect to Ramsay and Reynolds' (2000) development stages, it can be identified that some studies adhere to some of the stages, but rarely to all. The studies by Stevenson-Hinde and colleagues (Stevenson-Hinde and Zunz, 1978; Stevenson-Hinde *et al.*, 1980) do, however, appear to have followed the majority of such stages, thus adding further support to the adaptation of their assessment method for the measurement of horse personality.

2.4 Conclusion

In this review the theories, methods and reliability of human and animal personality research have been discussed. It is clear that human personality research has had some influence on animal personality research, although this is not always applied. The increasing volume of research on animals is beginning to provide strong evidence for the existence of animal personality. For the topic of horse personality the evidence is still growing, but has been able to demonstrate individual differences in horse behaviour.

Horse personality research has mainly been approached through the use of behavioural tests. These are, however, restricted in the breadth of information they can gather about an individual's personality as a whole. Furthermore they are often restricted to a short period of time and can be time consuming. An alternative approach to horse personality assessment may be that of trait rating. Animal studies using methodology similar to that of Stevenson-Hinde *et al.* (1980) use behaviourally defined adjectives or traits. Such terms have been transferred to other species (Caine *et al.*, 1983; Feaver *et al.*, 1986; Gold & Maple, 1994; Gosling, 1998; Wielebnowski, 1999; Martin, 2005), yet few, if any researchers have chosen to apply this method to horses. French (1993), Anderson *et al.* (1999) and Momozawa *et al.* (2003; 2005) used methods comparable to those of Stevenson-Hinde *et al.* (1980) but using different adjective lists. Other questionnaire studies on horses have tended towards using statements describing behaviour during specific events, for example, "*having feet picked up and trimmed*" (Seaman *et al.*, 2002), "*threatening towards unknown person*" (Momozawa *et al.*, 2003) and "*fearful when ridden*" (Le Scolan *et al.*, 1997). Finally, the use of a human based questionnaire by Morris *et al.* (2002a, 2002b) was seen as controversial due to its dependence on human personality

theory. The authors, however, did at least explore a comparative approach to horse personality research.

Previous horse personality assessment methods have been found to lack validity and reliability. There is a need, therefore, for the development of a horse personality assessment method that is both valid and reliable, but that can also be used for cross-species comparisons. Such an assessment method may also have practical applications within the equine industry. The Stevenson-Hinde *et al.* (1980) approach, however, has gone through appropriate development stages as described by Ramsay and Reynolds (2000). Furthermore it has been identified as being both valid and reliable when adapted for measuring personality across a range of animal species. Based on this the Stevenson-Hinde *et al.* (1980) rating method was selected for adaptation for the assessment of horse personality. Many of the adjectives used in the Stevenson-Hinde and Zunz (1980) adjective list were identified as being transferable to horses. This indicated that it would be suitable for the assessment of horses as well as non-human primates. This method was more favourable than the use of more human based assessment methods, as it allowed for flexibility when constructing personality dimensions. This in turn allowed for exploration of horse personality dimensions and for comparative research across species. This assessment method must, however, be tested for validity and reliability, by ensuring that all three of Gosling and Vazire's (2002) criteria are met.

3 Evaluation of a Novel Method of Horse Personality Assessment: Rater-Agreement and Links to Behaviour

3.1 Introduction

As identified in Section 2.3 the methods of horse personality that have been previously employed lack comparability to those of other species, thus limiting cross-species comparisons in personality and consistency in assessment methods. Furthermore, these studies have tended to focus on direct behavioural measurements (e.g. Le Scolan *et al.*, 1997; Wolff *et al.*, 1997; Visser *et al.*, 2001; Seaman *et al.*, 2002; Visser *et al.*, 2002; 2003a; 2003b) rather than the development of personality rating scales that have the potential to be applied quickly and easily by horse owners and handlers. In contrast some authors have developed novel rating systems and have utilised lists of either behavioural adjectives (Momozawa *et al.*, 2003, 2005) or have adapted human rating systems (Morris *et al.*, 2002a, 2002b). Such studies, however, have generally not identified any attempts to ensure rater or trait reliability. Furthermore, associations between personality ratings and recorded behaviour have been limited (Momozawa *et al.*, 2003). Previous studies, therefore, do not appear to have fully met Gosling and Vazire's (2002) criteria.

In contrast, Stevenson-Hinde and Zunz (1978) developed their trait list through communication with the regular handlers of the rhesus macaques used in their study and rigorously tested the reliability of their adjectives/traits. The original list developed in 1972 (see Stevenson-Hinde and Zunz, 1978) was developed from descriptors used by the regular observers of the macaques to describe their behaviour and characters. From these

descriptors 33 behaviourally defined adjectives were retained and used by the observers to rate each individual macaque's personality on a scale from one (extreme antithesis) to seven (extreme manifestation). The macaques were annually rated by handlers, between 1974 and 1977. The same three observers rated the macaques independently in both 1974 and 1975. Their ratings for each item were subsequently correlated across each monkey for both years in order to identify whether the macaques were being rated reliably. Following this analysis 19 items were retained having reached significant correlation between raters (Pearson product moment correlations, $P<0.05$ one tailed, r values not provided). Similar comparisons were made between two observers in 1976 using the 19 retained adjectives plus ten additional ones (29 in total). Of these 23 were identified as having been rated reliably ($P<0.05$, one tailed, r values not provided) and were utilised in the 1976 and 1977 assessments. Finally, after later analyses the list utilised by Stevenson-Hinde *et al.* (1980) incorporated a total of 25 adjectives/traits. This process helped to ensure a reliable list of traits that were easy to interpret and could be related to animals. The resulting data for each year were then entered into PCA to identify components/personality structures that were stable across years (Stevenson-Hinde *et al.*, 1980)

The rigorous manner in which Stevenson-Hinde and Zunz (1978) developed their rating scale meant that their rating data were shown to be reliable and appropriate for use by laypersons (animal handlers). Perhaps because of this stringency, applicability and ease of use, the final list has since been used as a template for several animal personality studies on a variety of animal species. In turn, this has allowed for easier cross-species comparisons in personality structure, some of which have been discussed by Gosling and John (1999).

When using the Stevenson-Hinde and Zunz (1978) trait list as a template, authors have adapted it to suit the animal species being studied. This has often resulted in both the removal of original traits and the addition of novel traits in order for the terms to be relevant to the lifestyle and behaviours of the species being studied (e.g. Caine *et al.*, 1983; Feaver *et al.*, 1986; Gold & Maple, 1994; Gosling, 1998; Wielebnowski, 1999; Martin, 2005). Although this may result in limitations during cross-species comparisons of personality structure, it is important that the rating criteria be relevant to the individual species and that assumptions are not made as to the structure of the resulting PCA.

This first experiment set out to meet two aims. The first aim was to adapt the Stevenson-Hinde *et al.* (1980) rating method for horse personality assessment by ensuring that the terms were relevant to horse personality. The adapted questionnaire was utilised to assess 61 horses and was then tested for reliability and validity using Gosling and Vazire's (2002) three criteria. In order to meet these criteria, multiple raters were used so that trait and rater reliability could be assessed (Criterion One) and personality data were correlated against behaviour data recorded whilst horses were at grass (Criterion Two and partially Criterion Three). The second aim was to use PCA to explore the resulting data for any underlying personality components, which were then compared to those identified in other species. These aims relate to General Aims I and II (Section 1.1) of the project.

3.2 Materials and methods

3.2.1 Experimental animals and management

A total of 61 horses kept at livery at four equine establishments in the UK were used in this study. All horses were mature, privately owned, well handled, in light exercise and from a variety of backgrounds and breeding. Horses were selected using two criteria. These were that study horses were: 1) regularly out at pasture during daylight hours and 2) kept in established social groups of a minimum size of two. The sample group contained 39 geldings and 22 mares. Ages ranged between 3 and 27 years and a mean age of 11.7 years (n = 58); the age of three horses was unknown. A total of 19 horses were pure bred (11 thoroughbreds, three warm bloods, two Arabs, one Friesian, one Welsh section C and one Irish Draught) with the remainder being either thoroughbred or Irish draught crosses (*n* = 22), or of unknown breeding (*n* = 20) (see Appendix 1). Husbandry procedures were similar across yards.

3.2.2 Personality assessment

The Horse Personality Questionnaire (HPQ) (Appendix 2) was constructed using 30 behaviourally defined adjectives (also referred to as traits) (Table 3.1). Of these, 25 were derived from the Stevenson-Hinde and Zunz (1978) and Stevenson-Hinde *et al.* (1980) Behaviour Rating Questionnaire and due to the questionnaire's previous adaptability to other species were considered to be suitable for the assessment of horses. Five additional adjectives were included to further adapt the questionnaire for horses. Three of these (suspicious, hard working and reliable) were adapted from Morris *et al.* (2002a). The final two (stubborn and intelligent) were commonly used by horse owners and handlers, and were considered to be suitable adjectives for the description of horse personality. In order

to ensure usability and applicability to horses the questionnaire was developed in communication with equine students and staff at Moulton College, Northampton, England.

Horses were scored on each trait using a seven point Likert-type scale (Reckase, 2000; Coolican, 2004) by asking respondents to circle the most appropriate number. A score of one represented no expression and a score of seven represented total expression, thus four represented an 'average' or intermediate score. The HPQ was a four-page questionnaire that included demographic questions about the horse and owner, as well as the 30 traits, each of which was accompanied by a full behavioural definition. Demographic questions were included in order to assess how long the handler had known the horse and to ensure the horses had remained at the same yard for at least the previous six months.

Each horse was assessed by three raters. Regular handlers of the horses completed the HPQs with each horse being assessed by at least two handlers. A regular handler was defined as someone who handled the horse at least four times a week and had been doing so for at least six months and included both owners and yard staff. In addition the author rated all horses on completion of behavioural observations (see Section 3.2.3) and is referred to in this study as rater one. Handlers were not constant across all horses; therefore for the purposes of data analysis, they were classified as either a rater two or rater three. Instructions on how to complete the questionnaire were provided and raters were instructed not to discuss their answers with each other.

Table 3.1: Behavioural definitions of personality adjectives (traits) used in the Horse Personality Questionnaire

Personality Adjective	Behavioural definition
[a]Active	Moves around a lot, does not like being still for long.
[a]Aggressive	Causes harm or potential harm to other individuals, both horse and human.
[a]Apprehensive	Seems to be anxious about everything, fears or avoids any kind of risk.
[a]Confident	Behaves in a positive, assured manner, not restrained, tentative.
[a]Curious	Readily explores new situations.
[a]Eccentric	Shows stereotypies, unusual mannerisms and exaggerated behaviour.
[a]Effective	Gets own way, can control others, fairly dominant individual.
[a]Equable	Reacts to others in an even, calm way; not easily disturbed.
[a]Excitable	Over reacts to any change, easily excited, highly strung.
[a]Fearful	Retreats readily from others or from outside disturbances.
[a]Insecure	Hesitates to act alone; seeks reassurance from others.
[a]Irritable	Reacts negatively with little provocation.
[a]Motherly	Provides warm receptive secure base for others, is tender and caring.
[a]Opportunistic	Seizes a chance as soon as it arises.
[a]Permissive	Could, but does not interfere with behaviour of others.
[a]Playful	Initiates play and joins in when play is solicited.
[a]Popular	Sought out as a companion by others.
[a]Protective	Prevents harm or possible harm to others.
[a]Slow	Moves and rests in a relaxed manner, moves slowly and deliberately, not easily hurried.
[a]Sociable	Seeks companionship of others.
[a]Solitary	Spends a lot of time alone by choice.
[a]Subordinate	Gives in readily to others, submits easily and does not put up a fight to defend self.
[a]Strong	Depends upon sturdiness and muscular strength.
[a]Tense	Shows restraint in posture and movement; carries the body stiffly, which suggests a shrinking tendency, as if to pull back and be less conspicuous.
[a]Understanding	Responds in a discriminating and appropriate manner to the behaviour of others.
[b]Suspicious	Doesn't trust others readily (human and horse), trusts few individuals.
[b]Reliable	Can be trusted to do things or behaves well, might also be considered a safe horse to be with.
[b]Hardworking	Keen to do well, behaves well during 'work', and concentrates on what it is being asked to do.
[c]Stubborn	Does not give in easily, not very cooperative.
[c]Intelligent	Learns new things easily/fast benefits from mental stimulation

[a] Adapted from Stevenson-Hinde *et al.* (1980); [b] Adjectives derived from Morris *et al.* (2002a); [c] Adjectives derived from behavioural terms

3.2.3 Behavioural observations

An ethogram (Appendix 3) was developed by combining behavioural definitions from five published works; Le Scolan *et al.* (1997), McDonnell and Haviland (1995), McDonnell and Poulin (2002), Strand *et al.* (2002) and McGreevy and Nicol (1998). It included measures of social behaviour, activity, aggression and feeding behaviour.

Behavioural observations were carried out on all 61 horses between July 2003 and October 2005 and between 11:00 and 15:00 hours on each observation day. A 30-minute acclimatisation period was allowed to pass before observations began, with the aim of reducing the observer effect (Martin & Bateson, 1993, p.31-32; Strand *et al.*, 2002). Behavioural data were collected using focal sampling and continuous recording methods (Martin & Bateson, 1993, p.84-85, 87-89) using a focal period of 15 minutes. Observation periods were short so as to reduce the effects of observer fatigue (Martin & Bateson, 1993, p.122). In total, two hours of observations (eight, 15-minute samples) were collected per horse (between 11:00 and 15:00) over a period of several days so as to provide an indication of the horses' typical behaviour whilst at grass. A compromise between the volume of observations and the overall sample size was necessary due to the time consuming nature of behaviour observations and the large sample size required to run the PCA. This meant that it was not practicable to collect additional hours of observations. Availability of the horses at the different yards also added further restrictions to the length of observations that were possible. The two hour total observation period provided sufficient data to allow for meaningful behaviour-personality correlations as Wielebnowski (1999) had previously demonstrated significant personality-behaviour correlations in cheetahs using much shorter periods of time (10 minutes).

For the purpose of behaviour recording, rater one was positioned outside the perimeter of the field containing the study horses so as to minimise any disturbance to the horses as well as to increase the safety of the observer. Due to the topography of the fields at yard 2, this was not possible, so the observations were carried out from a safe point within the field to allow for an unobstructed view of the horses. Behaviours were recorded at all sites, using a recording sheet, adapted from the ethogram, which measured both duration and frequency of behaviours. The horses were free to behave as normal and had access to the whole of their field. Interaction between rater one and the horses was kept to a minimum at all times.

3.2.4 Statistical analyses

Data were manipulated using Excel 2000 and statistical analysis was carried out using two statistical packages. The Kendall coefficient of concordance (W) (Siegel & Castellan, 1988) was calculated using Genstat 8th Edition for Windows 2000 (Lawes Agricultural Trust) and was employed to assess agreement between raters and trait reliability. These were also assessed using Spearman rank-order coefficients (r_s) (Siegel & Castellan, 1988) calculated using SPSS Version 13 for Windows 2000 (SPSS inc. US). Horse personality structure and its association with recorded behaviour were explored using PCA (Brace *et al.*, 2003) and Spearman rank-order coefficients (r_s) (Siegal & Castellan, 1988), both of which were performed using SPSS. Alpha was set at 0.05, as was utilised by Stevenson-Hinde and Zunz (1978) and Pederson *et al.* (2005) for comparable analyses.

Rater agreement and reliability of behaviourally defined adjectives

As the validity of the Stevenson-Hinde *et al.* (1980) questionnaire traits had already been demonstrated in a variety of species, the most likely source of unreliability in this data set was considered to be the raters. Their data were therefore tested for reliability prior to that of the traits.

In order to evaluate the reliability of the HPQ, rater agreement was tested. This was completed in two stages. The Kendall coefficient of concordance (*W*) was calculated for the trait scores of each horse between three raters. The Kendall coefficient of concordance is used to determine the association among *k* sets of rankings (Siegel & Castellan, 1988) and has been used to assess rater reliability in other animal personality studies (Wielebnowski, 1999; Morris *et al.*, 2002a; Martin, 2005). The use of Kendall's W helps to minimise the number of statistical tests carried out, thus reducing the occurrence of Type 1 error (i.e. rejecting the null hypothesis when it should be accepted). As in Martin (2005) any individual animals with personality ratings across raters failing to achieve a significant *W* value, were entered into Spearman rank-order correlations to identify if any rater pairs had agreed on the horses' scores. Any horses found not to have significant positive correlation ($P > 0.05$) between any two raters were removed from the next stage of analysis to test the reliability of traits. This was a modification of the technique employed by Feaver *et al.* (1986), to remove unreliable traits (i.e. those that raters fail to find significant agreement on for any of the individuals assessed) from further analysis and add rigour to the assessment method. For the purpose of this study it has been adapted for the removal of horses that raters could not find significant agreement on and were therefore perceived as being difficult to rate accurately.

Using data from reliable horses the reliability of traits was assessed. The r_s coefficient for each trait was calculated by comparing scores across horses, for example active was compared between raters for all horses. Any traits that did not have significant ($P > 0.05$, one tailed) positive correlation between any rater pairs were removed from further analysis and deemed to be unreliable as in Feaver *et al.* (1986). Such traits were interpreted as being unsuitable for the assessment of horse personality.

After the removal of unreliable traits rater reliability was re-assessed using the same method outlined above, thus *W* was re-calculated for all 61 horses. Spearman rank-order coefficients were calculated for any horses found not to have significant *W* coefficients ($P > 0.05$). Horses that did not have significant positive correlations ($P > 0.05$) between raters at this stage were removed from any further analyses due to the apparent difficulty in finding agreement on the scores of these horses. Such horses were removed as it could not be determined if their data were sufficiently accurate to provide a fair representation of their personality.

Horse personality structure

The mean trait scores were calculated for all horses previously found to have significant agreement between raters (after the removal of any unreliable traits). These were calculated using scores given by raters that had been shown to agree. Where only one pair of raters found significant agreement on a particular horse, ratings from the third rater were not used to calculate the mean score.

These scores were entered into a PCA with Varimax rotation (Brace *et al.*, 2003). Extracted components were determined using the eigenvalue criterion (i.e. extracted components must have an eigenvalue of greater than one, where the eigenvalue is a measure of variation explained by the component) and by using scree plots and the percentage variability accounted for by each of the components (Kline, 1993a; Brace *et al.*, 2003).

A scree plot shows the eigenvalue of each identified component. The shape of a scree plot is such that the slope is initially quite steep with the first few components having the largest eigenvalues (Kline, 1993a; Brace *et al.*, 2003). The point at which the slope begins

to level off (i.e. the eigenvalues of juxtaposed components become more similar) is a good measure of how many components should be included in the final model. If at this point, the component did not have a sufficiently high eigenvalue (greater than one) and was not explaining much of the total variance, then that component was rejected in order to simplify the overall model (Kline, 1993a; Brace *et al.*, 2003) and the PCA run again but fixed at the appropriate number of components.

A varimax rotation was used to calculate the simplest pattern of component loadings but with the maximum variation, making the resulting components easier to interpret in terms of personality structure (Kilne, 1993a; Brace *et al.*, 2003). Finally, Kline (1993a) recommended that a sample size of at least twice the amount of variables be used for PCA. As the HPQ contains thirty traits a sample size of 61 horses was deemed sufficient to allow for PCA. In order to test for stability of the mean ratings model, individual rater's data were entered into PCA separately using the same analysis technique and all models were compared.

Behaviour observations versus personality components

The final stage of data analysis explored associations between personality and observed behaviour. For the purpose of analysis the observed behaviours were combined to produce 13 behaviour categories (see Table 3.2), this is similar to the process carried out by Pederson *et al.* (2005) and aimed to reduce the number of correlations performed, thus limiting Type 1 error. Behaviours were grouped according to their similarities and their meaning. For example, kicking, kick threat, bite threat, biting, head threat given and nipping were grouped together as they can all imply an aggressive intention (Feh, 2005).

The mean duration (in seconds) or frequency of each of the behaviours was calculated for each of the horses that had been rated reliably. These values were then combined to produce the values for each behaviour category. Component scores for each horse were calculated automatically by SPSS as part of the PCA and were saved as variables. Component scores were calculated using an individual's score on each trait, the eigenvalue of the component and the trait loadings for that component (Kline, 1993a). The data for each of the behaviour categories were then entered into Spearman rank-order correlations with each horse's scores for each of the components identified by the PCA.

Table 3.2: Definitions of the 13 behaviour categories used for the behaviour-personality correlations. See Appendix 3 for full ethogram.

Behaviour category	Behaviours	Type of data collected	
		Frequency	Duration
Antagonised	Kicked	✓	
	Bitten	✓	
	Nipped	✓	
	Head threat received	✓	
	Submissive	✓	
Antagonistic	Kicking	✓	
	Biting	✓	
	Nipping	✓	
	Head threat given	✓	
Eat	Browse		✓
	Graze		✓
Exploration	Exploration		✓
Fast activity	Canter		✓
	Gallop		✓
Follow	Chasing		✓
	Herding		✓
Followed	Chased		✓
	Herded		✓
Groom	Mutual groom	✓	
	Self groom	✓	
	Roll	✓	
Idle	Standing		✓
	Laid down		✓
Play fight	Play fight		✓
Steady activity	Trot		✓
	Walk		✓
Stereotypy	Wind suck		✓
	Eat wood		✓
Vigilance	Vigilance		✓

3.2.5 Ethics and welfare

In order to ensure that the study was ethically sound and that animal welfare was not compromised, the methodology of this experiment was developed and assessed with reference to the Association for the Study of Animal Behaviour (ASAB) guidelines (ASAB, 2006). The final methodology was approved by the Moulton College Research Committee who acknowledged that, as no manipulation of the animals or their husbandry was involved, ethics and welfare were not compromised.

3.3 Results

3.3.1 Rater agreement and reliability of behaviourally defined adjectives

A total of 34 owners and handlers completed questionnaires for the study horses, with some raters assessing several horses. At stage one of the analysis, data for 35 of the 61 horses were found to have significant concordance ($n = 30$, $W \geq 0.467$, $P \leq 0.03$) between raters and were classified as reliable. Data for the remaining 26 horses were entered into Spearman rank-order correlations. Nine of the 26 horses showed significant correlation ($n = 30$, $r_s \geq 0.371$, $P \leq 0.044$) between at least one of their rater pairs. In total, 44 horses (72.13 %) were classified as reliable, having achieved either significant W or r_s coefficients, and were entered into the next level of analysis to test trait reliability. See Appendix 1 for W and r_s values for all 61 horses.

Spearman rank-order correlation coefficients were calculated to compare judge scores for each trait across all 44 reliably rated horses (Table 3.3). Where no significant correlation was shown, the data for that trait were classified as unreliable. Of the 30 traits analysed 25 had significant positive correlations ($n = 44$, $r_s \geq 0.329$, $P \leq 0.031$). A total of five traits (hard working, confident, permissive, solitary and strong) failed to show any significant positive correlation across any rater pairs and were removed from further analysis. It was also noted that agreement was highest between the regular handlers of the horses (raters two and three), with these raters only failing to find agreement on the five traits previously mentioned. In contrast the comparisons of rater one with raters two and/or three, only showed significant agreement on eleven of the traits (Table 3.3).

Rater agreement for all 61 horses was re-analysed after the removal of the five unreliable traits, using the same statistical procedure as at stage one of testing rater agreement. This

was carried out to assess whether the removal of unreliable traits would increase rater agreement. A total of 32 horses had significant concordance ($n = 25$, $W \geq 0.49$, $P \leq 0.023$) and were classified as reliable. Of the remaining 29 horses, 12 showed significant correlation ($n = 25$, $r_s \geq 0.403$, $P < 0.05$) between raters (Appendix 1). The remaining 17 horses showed no significant correlation between raters and were therefore classified as unreliable and their data removed from further analysis. This was necessary, as it could not be confirmed that their personality assessments were accurate and reliable, nor was it possible to determine which rater had provided the most accurate assessment. Of the 61 horses, data from 44 (72.1%) (32 geldings and 12 mares) were classified as reliable and entered into the PCA. In addition, as rater one was generally not shown to be in agreement with raters two and three, rater one's data were not used for the calculation of mean trait scores for entry into the PCA.

Table 3.3: Spearman rank order coefficients (r_s) of trait scores for 44 horses identified as having been rated reliably by their regular handlers.

Trait	r_s coefficients		
	Rater 1 vs Rater 2	**Rater 1 vs Rater 3**	**Rater 2 vs Rater 3**
Active	**0.52*** **	**0.51*** **	**0.76*** **
Aggressive	0.06	0.01	**0.61*** **
Apprehensive	0.24	0.06	**0.37***
Confident	0.19	0.14	0.30
Curious	0.03	0.20	**0.33***
Eccentric	0.09	0.01	**0.43****
Effective	**0.60*** **	**0.47****	**0.33***
Equable	0.09	0.15	**0.66*** **
Excitable	**0.33***	**0.36***	**0.76*** **
Fearful	**0.41****	**0.37***	**0.56*** **
Hardworking [a]			0.18
Insecure	0.05	0.11	**0.47****
Intelligent	0.29	0.15	**0.42****
Irritable	**0.28****	0.13	**0.54*** **
Motherly	**0.48****	**0.31***	**0.60*** **
Opportunistic	**0.35***	0.22	**0.33***
Permissive	-0.07	0.19	-0.07
Playful	**0.48****	**0.31***	**0.60*** **
Popular	-0.04	-0.01	**0.36***
Protective	0.20	0.26	**0.37***
Reliable	0.22	-0.03	**0.44****
Slow	**0.41****	**0.39****	**0.51*** **
Sociable	0.17	0.23	**0.60*** **
Solitary	0.07	0.02	0.24
Strong	0.21	0.08	0.24
Stubborn	0.04	-0.12	**0.61*** **
Subordinate	0.99	**0.30***	**0.42****
Suspicious	0.22	0.08	**0.43****
Tense	0.22	**0.51*** **	**0.38***
Understanding	-0.02	-0.06	**0.43****

[a] Rater one unable to score horses on hardworking. * $P<0.05$, ** $P<0.01$, *** $P<0.0001$.

3.3.2 Horse personality structure

The mean scores for each of the 25 traits were calculated for the 44 reliable horses using data from only raters two and three. These data were then entered into a PCA with varimax rotation. Seven components were originally extracted by the PCA, each with eigenvalues of greater than one and together explained 80.90 % of the total variance. Component Seven only accounted for 5.89 % of the total variance in the data and was rejected after examination of the scree plot (Figure 3.1) due to its low eigenvalue (1.47). In contrast, component six was retained as it had an eigenvalue of 1.66 and contributed 6.45 % of the total variance and was therefore thought to be a valid part of the overall model. The PCA was re-run to extract six components. The final model consisted of six components and explained 76.53 % of the total variance. The component structures, eigenvalues and item loadings for the six retained components are shown in Table 3.4. In addition, the component scores for each of the 44 horses are shown in Appendix 4.

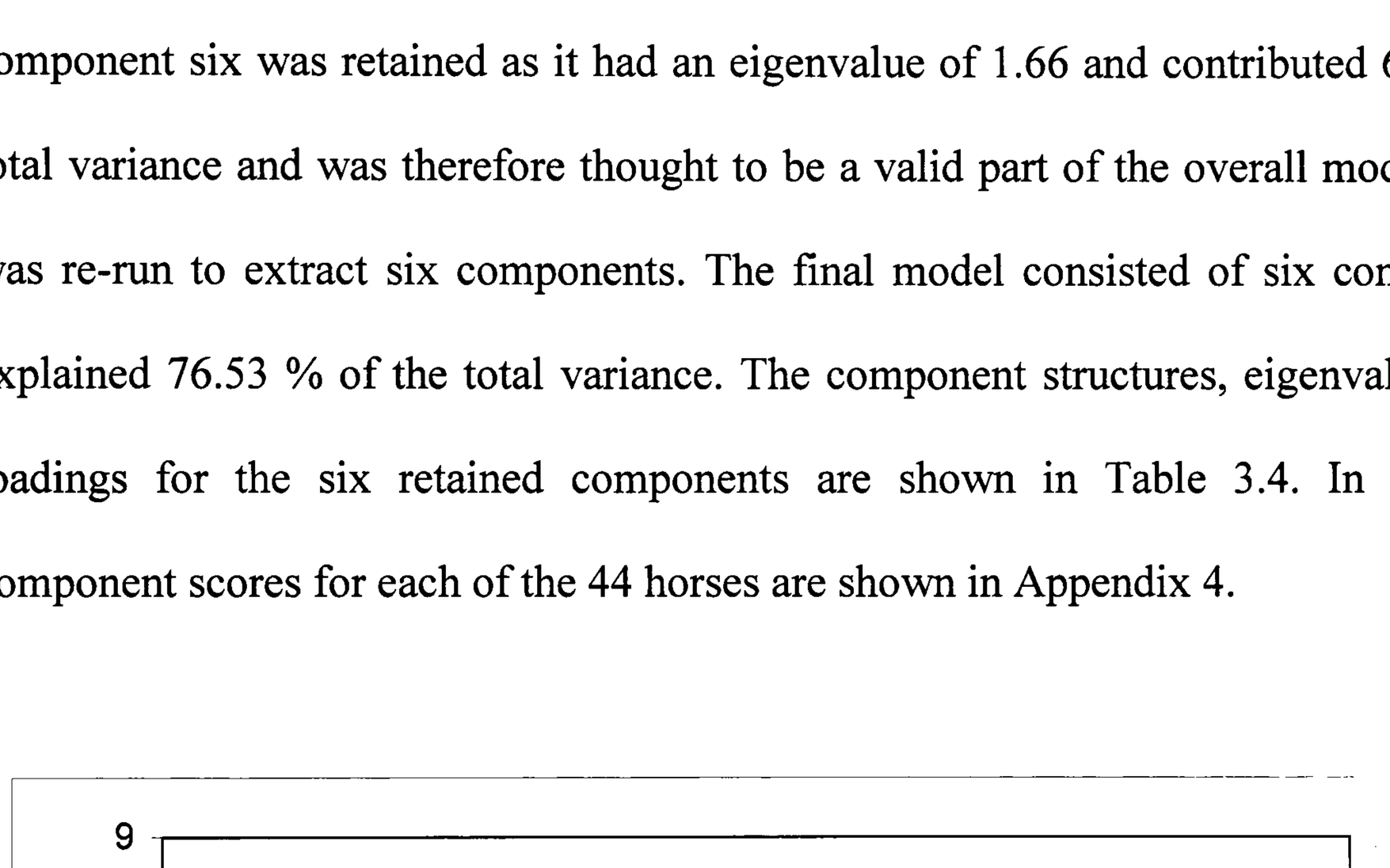


Figure 3.1: Scree plot showing the eigenvalues of each component extracted by principal component analysis, with varimax rotation, on horse personality questionnaire data for 44 horses and 25 traits

Table 3.4: Loadings of behaviourally defined adjectives (traits) onto six components extracted using principal component analysis with varimax rotation using data from 44 horses and 25 traits.

Traits	Component 1 Antagonism	Component 2 Anxiousness	Component 3 Activity	Component 4 Sociability	Component 5 Protection	Component 6 Inquisitiveness
Reliable	-0.792	-0.131	-0.17	0.015	0.168	-0.153
Equable	-0.721	-0.369	-0.115	-0.064	-0.032	0.315
Subordinate	-0.716	0.291	0.028	0.025	0.227	-0.097
Eccentric	0.718	0.212	0.165	0.004	0.293	0.163
Aggressive	0.719	-0.015	0.116	-0.236	-0.372	0.038
Stubborn	0.732	0.111	0.161	-0.141	-0.21	0.192
Irritable	0.862	0.257	0.172	0.028	0.012	0.006
Suspicious	-0.053	0.563	0.514	-0.344	-0.22	-0.255
Apprehensive	-0.003	0.887	0.217	0.061	-0.038	-0.041
Fearful	0.048	0.916	0.004	0.04	0.054	0.132
Insecure	0.175	0.828	-0.074	0.12	-0.045	-0.005
Tense	0.309	0.69	0.418	0.056	-0.054	0.207
Excitable	0.337	0.643	0.494	-0.013	0.143	0.088
Slow	-0.283	-0.306	-0.62	-0.362	0.022	0.067
Intelligent	0.046	0.193	0.752	-0.087	0.42	0.17
Active	0.144	0.359	0.656	0.369	0.038	0.317
Effective	0.333	-0.149	0.677	0.244	-0.108	0.14
Popular	-0.273	-0.105	0.058	0.765	0.359	-0.029
Sociable	-0.244	0.364	0.089	0.538	0.27	0.378
Playful	-0.003	0.224	0.203	0.821	0.125	0.06
Understanding	-0.566	0.041	0.094	0.172	0.705	-0.148
Motherly	-0.253	0.006	0.023	0.194	0.865	0.109
Protective	0.091	-0.174	0.06	0.537	0.707	-0.008
Curious	0.151	0.072	0.026	0.25	-0.007	0.762
Opportunistic	0.222	0.045	0.345	-0.26	0.035	0.626
Eigenvalue	5.03	4.40	2.90	2.59	2.58	1.64
Cumulative Variance (%)	20.12	37.70	49.29	59.66	69.98	76.53

Shaded cells represent the highest loading for each adjective across all components

The loadings of each trait were examined over the six components. The component where a trait had its highest loading was identified in order to determine which traits were providing the greatest contribution to each component. Each component was then interpreted by examining its contributing traits, and was given a descriptive label that provided an indication of the component's structure. A summary of the major contributing traits on each component is shown in Table 3.5. It can be seen from Table 3.5 that the first component was mainly comprised of those traits that might be associated with antagonistic behaviour and the agreeableness of the horse and was termed *Antagonism*. Component 2 was called *Anxiousness* and grouped together those traits that are associated with nervous behaviours. Component 3 comprised active and slow (-) as well as positive contributions from effective and intelligent and was termed *Activity*. The fourth component comprised interactive social behaviours and was termed *Sociability*. Protective traits were combined on Component 5 which was named *Protection.* Finally, Component 6 comprised only two traits and was named *Inquisitiveness*.

Table 3.5: Summary of component structures, listing the main contributing traits for each component.

Component 1 Antagonism	**Component 2 Anxiousness**	**Component 3 Activity**	**Component 4 Sociability**	**Component 5 Protection**	**Component 6 Inquisitiveness**
- Reliable	Suspicious	- Slow	Popular	Understanding	Curious
- Equable	Apprehensive	Intelligent	Sociable	Motherly	Opportunistic
- Subordinate	Fearful	Active	Playful	Protective	
Aggressive	Insecure	Effective			
Eccentric	Tense				
Irritable	Excitable				
Stubborn					

Comparison of PCA models

Rater 2 and Rater 3 data (44 horses) were entered into PCA (varimax rotation) separately in order to assess the stability of the model shown in Table 3.4 (Rater 1 data were not included having previously been identified as generally unreliable). The output of these analyses is shown in Appendices 5 and 6 (Rater 2) and 7 (Rater 3). The Rater 2 model originally extracted eight components which accounted for 81.35% of the variance (Appendices 5a and 5b). Upon reviewing the scree plot and eigen values, components 7 and 8 were rejected and a second was PCA run to extract six components. The new model accounted for 72.39% of the total variance (Appendix 6a and 6b).

Principal components analysis on the Rater 3 data extracted six components which accounted for 71.86 % of the variance (Appendix 7a and 7b). All three component structures (i.e. Rater 1, Rater 2 and mean of Raters 2+3) are compared in Table 3.6.

A total of 84% of the adjectives/traits were stable between the combined model and at least one other. It was, however, observed that components 3, 4 and 5 change order between the models. The association of particular adjectives/traits, however, remained similar across the different models. This switching of position was likely to have been due to the very similar eigenvalues of these three components within each model.

The model which used the mean values of Raters 2 and 3 was retained as the principal model and was used for subsequent analyses. This model was selected as it included data from both raters and accounted for the greatest amount of variance.

Table 3.6 Comparison of component structures between the three models extracted by PCA using individual data from Rater 2 and Rater 3 and the combined data of Raters 2 and 3.

Model (% variance)	**Component 1**	**Component 2**	**Component 3**	**Component 4**	**Component 5**	**Component 6**
Mean of Raters 2&3 (76.53%)	-Reliable ***	Suspicious**	- Slow***	Popular***	Understanding	Curious
	-Equable***	Apprehensive***	Intelligent***	Sociable**	Motherly	Opportunistic***
	-Subordinate***	Fearful***	Active**	Playful	Protective	
	Aggressive ***	Insecure***	Effective			
	Eccentric**	Tense**				
	Irritable ***	Excitable				
	Stubborn ***					
Eigenvalue	**5.03**	**4.40**	**2.90**	**2.59**	**2.58**	**1.64**
Rater 2 (72.39%)	Aggressive***	Apprehensive***	Active**	Motherly	Curious**	Effective**
	Eccentric**	Fearful***	Excitable**	Popular***	Playful**	Opportunistic***
	-Equable**	Insecure***	Intelligent***	Protective**		
	Irritable***	Suspicious**	-Slow***	Sociable**		
	-Reliable***	Tense**		understanding		
	Stubborn ***					
	-Subordinate***					
Eigenvalue	**6.18**	**4.69**	**1.94**	**2.43**	**1.48**	**1.38**
Rater 3 (71.86%)	-Aggressive***	Apprehensive***	Eccentric	Popular***	Curious**	Active
	-Irritable ***	Fearful***	-Equable	Protective**	Playful**	Effective**
	Motherly	Insecure***	Excitable**	-Suspicious	Sociable	Opportunistic***
	Reliable ***		Intelligent***			
	-Stubborn ***		-Slow ***			
	Subordinate ***		Tense			
	Understanding					
Eigenvalue	**6.43**	**4.38**	**3.16**	**1.55**	**1.42**	**1.03**

Traits were placed within the component where the trait had its highest loading.* indicates that a trait loaded onto the same component on two (**) or three (***) of the PCA structures calculated.

3.3.3 Behaviour observations versus personality components

Spearman rank-order correlations were carried out between the 13 behaviour categories and each of the six principal components. Only eight significant correlations were identified between behaviour categories and component scores (n = 44, $r_s > 0.32$, $P < 0.05$) and these are summarised in Table 3.7.

Table 3.7: Spearman rank order correlations between personality component scores for the 44 reliably rated horses, against corresponding behaviour data.

Behaviour Category	Components					
	Antag.	**Anx.**	**Activ.**	**Soc.**	**Prot.**	**Inquis.**
Antagonised	-0.18	0.01	-0.17	-0.30	0.15	0.18
Antagonistic	0.32*	-0.19	-0.18	-0.14	0.38*	-0.12
Eat	-0.10	0.07	0.23	0.17	-0.05	-0.42**
Exploration	0.04	0.13	-0.15	0.12	-0.06	0.39**
Fast activity	0.14	0.04	0.19	0.00	0.41**	0.21
Follow	0.28	-0.10	0.06	-0.10	-0.30	-0.03
Followed	-0.08	-0.03	-0.21	-0.19	-0.27	0.32*
Groom	0.02	-0.09	-0.18	-11	0.26	0.07
Idle	0.14	-0.12	-0.43**	-0.12	-0.13	0.30
Play fight	-0.16	0.02	-0.05	-0.36*	0.02	0.05
Steady activity	0.20	-0.15	-0.18	0.01	-0.07	0.10
Stereotypy	0.01	0.08	0.07	0.15	0.04	0.02
Vigilance	-0.17	0.14	0.18	0.07	0.27	0.03

* $P<0.05$; ** $P<0.01$. Antag = *Antagonisitic*; Anx. =*Anxiousness*; Activ. = *Activity;* Soc. = *Sociability;* Prot. = *Protection*; Inquis. = *Inquisitive*.

3.4 Discussion

The results demonstrated that the HPQ was both a reliable and valid method of personality assessment and six-component structure of horse personality. The adapted version of the Stevenson-Hinde *et al.* (1980) questionnaire was a reliable method of personality assessment, with only four of the original traits and one of the added traits found to be difficult to assess. Personality component scores were also found to significantly correlate with some of the observed behaviours recorded whilst the horses were at grass.

3.4.1 Rater agreement and reliability of behaviourally defined adjectives

Reliability between raters was high, with scores for 44 of the 61 horses (72.1%) being agreed on. This high level of agreement was the same as that found in the study by Martin (2005) on chimpanzee personality, who found agreement for 72.1% of the chimpanzees studied. Rater reliability for horse personality assessment was also tested by Morris *et al.* (2002a). Nine raters assessed ten horses using the NEO-PI-FFI (e.g. Costa & McCrae, 1992a) and the correlations between the horses' scores on each of the five factors indicated good levels ($r_s > 0.37$) of agreement between raters for three out of the five factors. Morris *et al.* (2002a) did not, however, clearly compare the scores given by raters for each horse, as has been done for the present study, and cannot therefore be directly compared. In contrast Anderson *et al.* (1999) failed to achieve high levels of reliability between raters of horses with agreement between rater pairs varying between 29% and 50% of horses at each yard, however no information on overall reliability was provided. Some horse personality studies do not appear to test rater reliability, for example Momozawa *et al.* (2005) used a 20-item questionnaire to assess the personality of 139 horses over two study periods. Although three different handlers rated each horse, the authors did not discuss tests of reliability or agreement between raters.

After the removal of the five unreliable traits significant W values ranged from 0.49 to 0.78 and significant r_s values from 0.40 to 0.66 for associations between raters. These coefficients were comparable not only to other animal personality studies, which on average reached coefficients of 0.52 (Gosling & Vazire, 2002), but also to the average correlation coefficient (0.50) found in human studies (Funder, 1995). The results of this study were, therefore, at a level of agreement and reliability that is not only accepted by animal personality research, but by that of humans. The large number of correlation analyses carried out in order to identify rater agreement within this study may have increased the possibility of Type 1 error. The level of alpha used was, however, comparable to that used in other animal personality studies (e.g. Stevenson-Hinde and Zunz, 1978; Stevenson-Hinde *et al.*, 1980; Pederson, *et al.*, 2005) and was therefore deemed to be sufficient and the results are thought to have provided an accurate representation of the reliability of horse personality assessment. In demonstrating the reliability of these data these results have provided further support for the ability of human raters to detect differences in horse personality.

The high level of agreement demonstrated in the present study provided evidence that these data met Gosling and Vazire's (2002) Criterion One that independent assessments must agree. There was, however, still a proportion of error between raters both in this study and those of other animal personality studies (see Gosling & Vazire, 2002), hence there is a need for animal personality researchers to explore the possible causes of error, so that it can be minimised in future studies (Gosling & Vazire, 2002). Possible causes may include the duration or type of acquaintance between raters and horses, the developmental stage of the animals or the characteristics being assessed (Gosling & Vazire, 2002). In the present study it was found that those people that regularly handled the horses (raters two and three) were much more likely to agree on the personality scores of the horses. In contrast rater

one was far less able to reach agreement with other raters and could be described as a 'bad judge' in relation to the categories of Funder *et al.* (1995). This may have been because she had known the horses for a shorter period of time than the handlers and had only observed the horses whilst at grass. It was also possible that some individuals were less consistent in their behaviour and as such could be described as 'bad targets' for personality assessment (Funder *et al.*, 1995).

The raters used in this study were shown to be competent at using the HPQ and only failed to find significant agreement on five of the 30 traits (hard working, confident, permissive, solitary and strong). With the exception of hard-working, all of these traits originated from the Stevenson-Hinde *et al.* (1980) rhesus macaques rating questionnaire. It was possible that the rejected items were more difficult for raters to interpret either due to lack of familiarity with the terms and their meaning, or lack of observation of the horses over a variety of situations. Future development of the HPQ will need to take into account such issues and re-assess the use of these terms. The process of assessing the reliability and practicality of traits is an important part of developing personality assessment methods. For example, although the original version of the Stevenson-Hinde and Zunz (1978) questionnaire contained 33 adjectives, several were removed and replaced as a consequence of reliability analyses, resulting in the final list containing 25 adjectives (Stevenson-Hinde *et al.*, 1980).

High levels of agreement between raters meant that only 17 horses were not entered into the PCA. Such data were not included as the lack of agreement made the data unreliable and provided unclear representation of the horses' personality and as such would have skewed the PCA results and given a false interpretation of horse personality structure.

The results of the present study demonstrated the ability of handlers to rate horses reliably as well as demonstrating the suitability of the HPQ to assess horse personality. Further research, however, is required to explore the reasons that some horses are more difficult to rate than others.

3.4.2 Horse personality structure

The PCA revealed six personality components, which were interpreted as being dimensions of horse personality. The total variance (76.53%) explained by all six components was comparable to that found in other animal personality studies (King & Figueredo, 1997, 72.4%; Gosling, 1998, 75%; Momozawa *et al.*, 2003, 71.4%; Martin, 2005, 78%; Momozawa *et al.*, 2005, 84%), and was higher than that found in others (Stevenson-Hinde & Zunz, 1978, 60%; Stevenson-Hinde *et al.*, 1980, 66-69%; Murray, 1998, 57.5%; Morris *et al.*, 2002b, 41.5%; Ley *et al.*, in press, 32.6%; McGrogan *et al.*, in press, 59%). Due to the necessity of removing unreliable horses, the sample size used for the PCA was slightly lower than that recommended by Kline (1993a) who suggested using twice as many individuals as variables. The final sample size was, however, comparable to that used by Stevenson-Hinde and Zunz (1978) who performed PCA on data from 46 animals and 23 items and also Stevenson-Hinde *et al.* (1980) who used 46 animals and 21 items. Furthermore, when PCA was carried out on the individual raters data the extracted components showed a high level of stability between models with few traits switching between components. It was therefore concluded that the limited sample size did not significantly affect the PCA. As such, the resulting component structure may provide a foundation for further horse personality research.

The use of PCA in equine personality research is still fairly recent and has not been applied in many equine personality studies for the analysis of questionnaire data. One example of

PCA being used for this purpose is the study on donkey personality by French (1993) who used PCA on data from 45 donkeys rated on eight paired behavioural adjectives. Principal components analysis extracted two components, *Obduracy* and *Vivacity*. The small number of variables may have limited the extraction of components. In comparison, the present study employed a more comprehensive adjective list, previously tested on a variety of other mammals (Stevenson-Hinde & Zunz, 1978; Stevenson-Hinde *et al.*, 1980; Feaver *et al.*, 1986; Gold & Maple, 1994; McGuire *et al.*, 1994; Wielebnowski, 1999; Martin, 2005) and extracted six components to provide a detailed representation of horse personality structure.

Morris *et al.* (2002a, 2002b) also demonstrated a multi-factor structure of horse personality. Horses were assessed using a human personality assessment method (the NEO-PI-FFI) (see Costa & McCrae, 1992a, 1992b), which is based on a five-factor model *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness* and *Neuroticism*. The horse personality data were then applied to these factors/components. The horse data fitted well to three of the factors (*Conscientiousness*, *Extraversion*, and *Neuroticism*) but raters found it hard to apply *Agreeableness* and *Openness* to horses. Low agreement on the *Agreeableness* factor was linked to the relationship between rater and horse, it was suggested that it involved a more personal judgement as to whether an individual found a horse agreeable or not. Low agreement on the *Openness* factor may be due to difficulty in assessing this scale on horses using an assessment tool designed for humans. Many personality terms can be readily applied to both humans and horses. Those in the *Openness* factor, however, are more specifically tailored towards humans, and include statements that are difficult, if not impossible, to apply to horses. Examples include "*has day dreams but does not like day-dreaming*" and "*gets enchanted by the natural world around him/her*". Although Morris *et al.* (2002a) aimed to test rater reliability; their assessment method had

not previously been validated for use on horses and highlights the importance of trait selection and reliability testing.

Despite having adapted and used a trait list originally designed for rhesus macaques, the components extracted by the PCA were comparable to those found in other horse studies. Momozawa *et al.* (2005) identified a four-factor structure of horse personality (*Anxiety, Trainability, Affability* and *Gate entrance*) and Creighton (2003) identified a five-factor structure (*Extraversion, Agreeableness, Emotionality, Openness* and *Conscientiousness*). The horse personality taxonomy identified in this study seemed comparable to that extracted by Creighton (2003), but with an additional component. Both models included elements of sociability (*Agreeableness vs Sociability*), anxiety (*Emotionality vs Anxiousness* and/or *Activity*) and curiosity (*Openness to experience vs Inquisitiveness*).

The personality structure demonstrated in the present study also showed some similarities to the Big Five used by Morris *et al.* (2002a) although the models do not appear to be identical. Firstly, the components *Neuroticism* and *Anxiousness* both contain items relating to fearfulness, tenseness and insecurity. *Anxiousness* did not, however, contain items relating to sadness, depression, helplessness, shame and the need for support from others unlike *Neuroticism*. Similarly, *Extraversion* appeared to be comparable to two of the horse personality components; *Sociability* and *Activity*. These components contained items relating to interaction with other individuals and high levels of activity. *Extraversion*, however, contained several items that have not been explored within this study, such as 'has a good sense of humour', 'light hearted', 'is an optimist' and 'prefers to be on his/her own' and is not, therefore a perfect match to these components.

Some similarities between *Agreeableness* and Antagonism were also identified. *Agreeableness* contained a mixture of both negative and positive social behaviours, with opposing contributions (i.e. negative/positive) as were seen in *Antagonism*. The traits included in *Agreeableness* also included elements of calculative behaviour and insight, such as 'feels others will take advantage if they can' and 'is selfish and egotistical'. Such terms were considered difficult to assess in horses and were not included in this study. In addition, there were limited similarities when *Openness* and *Inquisitiveness* were compared, although both components contained an element of curiosity and *inquisitiveness*. As previously mentioned in Section 2.1.3, the assessment terms used by Morris *et al*. (2002a) to assess *Openness* included several terms with an emphasis on insight and the 'thoughts' of the horses. These were identified as being difficult for raters to assess. Similar terms were not included within this study so as to increase rater reliability and ensure that the traits were applicable to horses. Finally in the comparison of *Conscientiousness* with *Protection* few similarities were identified. Both implied levels of dependability, however, many of the items included in *Conscientiousness* were not used within this study. Furthermore, within this study, the trait 'reliable' loaded most strongly onto *Antagonism* whereas it is normally included within *Conscientiousness*.

Although the personality components identified within this study did not match perfectly with other models of horse personality, they did show sufficient similarities to indicate that they provided a detailed description of horse personality. As the PCA was not forced to fit the data to a specified model, such as the Big Five, the resulting structure was more likely to be a true representation of horse personality. Furthermore, some of the components identified have been shown to be recurrent in other animal personality studies (Gosling and John, 1999).

Personality taxonomy: Cross species comparison

Several of the components identified within this study bear similarities to those components identified in other animal species. In a cross species review of 19 factor analytic studies of 12 species, Gosling and John (1999) compared animal personality taxonomies to the Big Five. Those components relating to *Extraversion, Neuroticism* and *Agreeableness* were identified as being recurrent in most of the species reviewed. Although the component labels differed, their comprising traits were very similar. Elements of *Openness* were identified in seven of the 12 species, and tended to include traits such as curiosity and exploration, some studies however, failed to include traits that linked to this component. *Conscientiousness* was only identified in the chimpanzee and human studies, but has since been identified in horses through the use of a 16-item rating questionnaire and PCA (Creighton, 2003). Nevertheless, it was not identified in this study, perhaps as a result of a lack of traits relevant to *Conscientiousness* (e.g. neat and clean, organised, methodical and systematic, Morris *et al.*, 2002a). In contrast the component *Antagonism/Dominance* is unique to animal personality and is not generally identified in humans, perhaps as a result of the complicated multiple dominance hierarchy systems present in human society (Gosling & John, 1999). In animals it has been identified as a separate personality factor in several species including chimpanzees (King & Figueredo, 1997), gorillas (Gold & Maple, 1994), rhesus macaques (Stevenson-Hinde & Zunz, 1978; Stevenson-Hinde *et al.*, 1980) and spotted hyenas (Gosling, 1998). The taxonomy of horse personality described in this study, shows similarity to the Big Five, and the factor structures found in other horse personality studies, thus providing further evidence for the efficacy of this assessment method.

3.4.3 Behaviour observations versus personality components

The present study set out to demonstrate a link between rated personality and measured behaviour. Significant correlations were identified between eight behaviour categories and five of the personality components. These correlations provided an encouraging link between personality and behaviour. These correlations are discussed below and compared to those found in similar studies.

The component *Antagonism* was positively correlated with antagonistic behaviours, indicating that an individual scoring highly on this component would perform more aggressive behaviours such as kicking, biting, and giving head threats, as would be expected. *Activity* was shown to negatively correlate with the behaviour category 'idle' this included standing and lying down and was interpreted to mean that horses scoring highly on *Activity* being more active and standing less compared to those horses that had lower scores. *Sociability* was positively correlated with antagonistic behaviours and fast activity (i.e. cantering and galloping) behaviours. These categories were often recorded during positive social interactions such as play fighting. It therefore seemed unusual that play fighting should be negatively correlated with *Sociability*. It is possible that this may have been as a result of observer error. Play fighting and 'real' fighting can be difficult to differentiate when being observed. If these instances were wrongly classified by raters two and three or by rater one during behaviour observations, then this could explain the negative correlation. Future work may benefit from exploring this relationship further.

The fifth component *Protection* was positively associated with antagonistic behaviours, follow and fast activity. These are not behaviours that may first be expected to be associated with protection, yet these behaviours are associated with defence and therefore

the protection of an individual/object. These behaviours do not, however, relate to the 'understanding' element of *Protection*.

Inquisitiveness was correlated with explorative behaviour but was also negatively correlated with eating behaviours, indicating that horses that were more explorative spent less time eating and presumably more time exploring their environment. Finally, none of the recorded behaviours correlated with *Anxiousness*. This may have been due to the horses being observed in a familiar environment and therefore showing fewer signs of anxiousness and fear. Such behaviours may be more readily observed in a novel or changing environment.

The r_s coefficients between the behaviour categories and personality components suggested moderate relationships between personality and behaviour and were comparable to those found by Pederson *et al*. (2005). The relationships identified provided evidence that the scores given to the horses by familiar raters reflected the true personality of the horses and were not just the implicit personality theories of the raters (Kenrick & Funder, 1988; Gosling & Vazire, 2002). They also suggest that ratings are a reflection of real behaviours (Kenrick & Funder, 1988; Gosling & Vazire, 2002). Although this study has not attempted to predict behaviour using the results of the HPQ, the demonstrated links between personality and behaviour indicate that such predictions may be possible and will be further explored in Chapter 5. As such these data have met Criterion Three ("*ratings must reflect attributes of targets, not observer's implicit personality theories*") and suggest the future fulfilment of Criterion Two ("*assessments must predict behaviours and real-world outcomes*") as proposed by Gosling and Vazire (2002, p.608). In turn, they add to the ever-increasing evidence for the existence of horse personality and our ability to measure it accurately.

Previous studies that have attempted to directly link behaviour and animal personality have so far provided mixed results with varying levels of success. Visser *et al.* (2003a) used behavioural tests (novel object, handling test, reward-learning test and avoidance-learning test) on 41 horses to measure specific aspects of personality; the same horses were then assessed on jumping ability. These data were combined with the behaviour test data and entered into a PCA to produce two performance components accounting for 64.8% of the total variance. The first component (PerfC1) was used as an integrative performance variable and entered into a multiple regression against behavioural variables measured during training sessions. The results were unclear, but indicated an association between specific behavioural variables and the personality of the horses.

In contrast, Pederson *et al.* (2005) successfully demonstrated significant associations between personality scores and observed behaviour in chimpanzees. Personality ratings were collected using an assessment method developed in a previous study (King & Figueredo, 1997) that had shown a six-factor structure of chimpanzee personality (*Extraversion*, *Dependability*, *Agreeableness*, *Emotionality*, *Openness* and *Dominance*). Behaviour observations and ratings were collected for 47 chimpanzees and compared using Spearman rank order correlations. Significant correlations ($P < 0.05$) were found for 14 component-behaviour pairs. Spearman coefficients were low ($r_s \leq +/- 0.29$) suggesting only moderate relationships, but still provided an important starting point for future research exploring the links between personality and behaviour.

3.5 Summary

Previous attempts at horse personality assessment have been inconsistent with studies assessing personality in other animals. Methodology has included behavioural assessment and the creation of novel rating methods. The present study has demonstrated that an assessment method originally developed for rhesus macaques and previously shown to be reliable for a variety of different species can also be applied to horses. Rater reliability was high, with raters being able to agree on the scores of 72.1% of horses with only five behaviourally defined adjectives being removed due to lack of rater agreement. Resulting PCA analysis revealed a six-component structure, the components of which were significantly correlated with 20 different recorded behaviours. Due to the strong links between personality and behaviour, it may be possible to test the ability of using personality to predict future behaviour in horses. Further development and adaptation of the HPQ may enable it to be implemented in the selection of horses for specific equine disciplines such as show jumping or dressage. The selection of horses with the most appropriate personality type for a specific discipline should reduce costs for trainers and improve horse welfare by avoiding the training of inappropriate horses.

This experiment has successfully met general Aims I and II of the project (Section 1.1) by demonstrating a reliable method of personality assessment and the existence of six horse personality components. The application of the HPQ to demonstrate further real-world outcomes, such as breed differences in behaviour, would further add to the validity of this assessment method.

4 Horse Personality: Variation Between and Within Breeds

4.1 Introduction

In Chapter 3, the HPQ was identified as a reliable and valid method of horse personality assessment. In this chapter the ability of the HPQ to be used on a larger scale and explore the existence of breed typical personalities in horses was assessed. Horse breeds are often described as having 'breed typical behaviours' and such claims are supported by anecdotal evidence from breed enthusiasts, with societies often promoting a breed by describing its typical temperament and personality. For example, the highland pony is described as having a "*kindly nature and even temperament*" (Highland Pony Society, 2006) and the Irish draught horse is described as having "*an intelligent and gentle nature and is noted for its docility and sense*" (Irish Draught Horse Society, 2006). In contrast, the Arab has been described as "*spirited, enduring, intelligent, bold, perceptive, sensitive and thoughtful*" (Foster, 2005), but they have also been known for being difficult when asked to do something against their will (Foster, 2005). There is growing scientific evidence that such breed typical personalities in animals do exist, it is therefore a good test of validity to use the HPQ to explore such differences.

Furthermore, it is now well accepted that behaviour and personality are controlled to some extent by genetics (Bouchard & Loehlin, 2001; Reif & Lesch, 2003; Mormède, 2005; van Oers *et al.*, 2005). Current research is beginning to quantify the heritability of specific behaviours and personality constructs, for example, *Dominance* in chimpanzees (Weiss *et al.*, 2000; Weiss *et al.*, 2002), exploratory and other behaviours in great tits (Dingemanse

et al., 2002; van Oers *et al.*, 2004; Carere *et al.*, 2005), *Dominance* in dogs (Svartberg, 2005; Pérez-Guisado *et al.*, 2006) and temperament in cattle (Gauly *et al.*, 2001). The genetic control of behaviour suggests that artificial selection can be used to select for desired behaviour types and is therefore likely to have resulted in breed typical personalities (Svartberg, 2006).

Breed differences in horse personality have not yet been directly explored, although some studies have looked at breed effects on behaviour. For example Hausberger *et al.* (2004) looked at the reaction of horses, from 16 breeds, to a bridge test and identified breed differences in the length of time taken to cross the bridge. Hausberger and Muller (2002) also found variation in friendly behaviour and reactivity between thoroughbreds, French saddlebreds and Angloarabs. Breed effects have also been identified in stereotypic behaviours (Luescher *et al.*, 1998; Redbo *et al.*, 1998; Houpt & Kusunose, 2000), in that thoroughbred horses are generally more susceptible to stereotypies, such as crib-biting and weaving, than are other breeds.

Breed differences in personality and behaviour have, however, been studied in much greater detail in dogs. For example, using behaviour data from over 13,000 Swedish dogs from 31 breeds, Svartberg (2006) demonstrated that dog breeds vary significantly on the traits playfulness, curiosity/fearlessness, sociability and aggressiveness. Svartberg (2006) also demonstrated that the most recent selection pressures (mainly breeding for the show ring) had a significant effect on current breed-typical behaviours. Notari and Goodwin (in press) and Bradshaw *et al.* (1996) were also able to identify breed differences in behaviour. They asked veterinarians and dog-care professionals to rank dog breeds on a selection of behavioural traits using their own experiences. Evidence of breed differences in dogs provides support for the hypothesis that horse breeds differ in personality.

It has also been identified that the temperament, and therefore personality, of a horse is considered to be an important attribute to horse owners and is considered a key issue in horse health and performance (Buckley *et al.*, 2004). Therefore a greater understanding of the typical behaviour and personality of specific horse breeds may aid the selection of horses for specific equine disciplines, including use for leisure by amateur riders. More informed selection of horses may lead to improved horse welfare, as horses could then be selected for appropriate functions and rider capabilities (Visser, 2003a). Furthermore, if breed typical personalities are identified and support the anecdotal evidence (a real-world outcome), then this will provide further evidence that the HPQ meets Gosling and Vazire's (2002) Criterion Two (assessments must predict behaviour and real-world outcomes).

This chapter focuses on the third general aim of the project (Section 1.1) and explores the potential differences in personality between eight different horse breeds using the HPQ developed in Chapter 3. The results of this survey will have both practical implications for horse management as well as implications for personality research.

4.2 Materials and methods

Personality data from 1223 horses from eight different horse breeds were collected, using the HPQ, and were compared. The breeds selected were: Irish draught horses, thoroughbreds, Shetland ponies, Arabs, Highland ponies, Welsh ponies and cobs, American quarter horses and Appaloosas. The breeds were selected to represent a variety of types that included lightweight, draught and pony breeds.

4.2.1 Questionnaire design

Horses were assessed using the HPQ (Appendix 8) as developed in Chapter 3, but with only 25 traits, having removed the five unreliable ones identified in Section 3.3.1. Personality assessment was carried out as described in Section 3.2.2 although each horse was only assessed by one regular handler. The data were transformed into scores for each of the six horse personality components (*Antagonism, Anxiousness, Activity, Sociability, Protection* and *Inquisitiveness*) as described in Section 4.2.3. The designation of the traits to their relevant components is shown in Tables 3.4 and 3.5.

The HPQ included a worked example and instructions on how to complete the questionnaire. Demographic questions about the focal horse and the person completing the questionnaire were also included in order to gather information on age, gender, breed and also the duration of the human-horse relationship. These questions enabled inappropriate questionnaires to be removed from analysis (i.e. where horses were too young, an inappropriate breed or had not been kept by the owner for long enough etc.). The same questionnaire format was used for all breeds.

4.2.2 Questionnaire distribution

Data collection was ongoing between January 2005 and January 2006. The HPQ was distributed in a variety of ways. Only pure bred horses were required for the study, therefore, a targeted sampling approach was used (Coolican, 2004), such that only owners/handlers of pure bred horses were asked to be involved. Approximately 4000 questionnaires were distributed directly to owners of pure-bred horses either through the relevant breed societies or by direct communication from the author. The project was also publicised in the newsletters of the relevant breed societies (see Table 4.1). The project and the HPQ were also advertised using articles in the local (Northamptonshire, UK) and British equine press. Respondents were invited to be involved with the project by either completing a paper version of the questionnaire or by completing an online version, which was hosted by Harper Adams University College and was available via a link on the Moulton College website from May 2005 to December 2005. See Table 4.1 for the societies involved, number of questionnaires received and the age and gender details for each breed used in this study. Data from mares, stallions and geldings were combined for analysis.

Data were included in analysis if the following criteria were met; 1) horses were a minimum of one year of age to allow for some stability in the personality (Visser 2001); 2) respondents had known the horse for a minimum of six months and regularly handled the horse being assessed; and 3) horses were pure-bred.

Table 4.1: Demographic and society information for each of the eight horse breeds assessed in the study

Breed	Society	Horse gender			Mean horse age (± SD)	HPQ		Total HPQ
		Gelding	Stallion	Mare		Online	Paper	
IDH	Irish Draught Horse society	50	6	94	10.5 (± 4.9)	23	127	150
TB	*n/a*	179	4	98	13.0 (± 6.2)	254	27	281
Shet.	Shetland Pony Studbook Society	53	17	88	11.5 (± 7.0)	34	124	158
Arab	Arab Horse Society	150	6	119	13.3 (± 6.7)	267	8	275
High.	Highland Pony Society	23	2	36	10.6 (± 5.1)	31	30	61
WPC	Welsh Pony and Cob Society	58	3	36	12.6 (± 6.3)	86	11	97
App.	British Appaloosa Society	66	2	62	12.0 (± 7.3)	130	0	130
AQH	*n/a*	42	1	28	15.0 (± 7.8)	71	0	71
Total		**621**	**41**	**561**	**12.4 (±6.6)**	**896**	**327**	**1223**

Gelding = castrated male; Stallion = intact male, Mare = female, SD = Standard deviation, *n/a* = Not applicable; IDH = Irish draught horse; TB = thoroughbred; Shet. = Shetland pony; High. = Highland pony; WPC = Welsh ponies and cobs; App. = Appaloosa; AQH = American quarter horse; HPQ = Horse Personality Questionnaire.

4.2.3 Calculation of component scores

For each breed the component scores for each horse were calculated using the horse personality model produced in Chapter 3 (Table 3.5). Component scores were calculated using Microsoft Excel in order to use the same component structure calculated in Chapter 3. An individual's score for each component was calculated by: (the individual's rating on an item) multiplied by (the loading of that item divided by the eigenvalue of the component) summed over all trait items. The loading is the calculated input of a trait onto a component and the eigenvalue is the variance accounted for by that component, as calculated during the original PCA shown in Chapter 3 (the designation of the traits to their relevant components is shown in Table 3.5). This is an adaptation of the equation used by Stevenson Hinde *et al.* (1980) and Stevenson-Hinde and Zunz (1978), but does not standardise the rating data prior to calculation of the component scores. Data were not standardised, as all variables were measured on the same scale and therefore did not require balancing for the effects of different measurement scales.

4.2.4 Statistical analyses

Questionnaire data were manipulated using Excel 2003 (Microsoft) and were analysed using Minitab 13 for Windows and SPSS Version 14 for Windows (SPSS Inc., USA). The aim of the statistical analysis was to explore any differences in component scores between breeds. The value of alpha was set at 0.05 for all statistical tests.

The component scores were compared across all breeds using the Kruskal-Wallis test (Siegel & Castellan, 1988). Where significant differences across breeds were identified ($P<0.05$), post hoc multiple comparisons tests were then performed (Siegel & Castellan, 1988) on all possible breed pairs (28 in total) to explore specific breed differences. Multiple comparisons tests (*a posteriori* tests) are commonly performed when a multi-

sample hypothesis of equal means/medians has been rejected, in order to identify where the differences between treatments/groups lie (Zar, 1999). Such tests are more reliable and valid than the use of multiple t-tests/Mann-Whitney U tests and reduce the chance of Type I errors as differences are identified in one test (Zar, 1999).

Similarly, the Kruskal-Wallis test was carried out to test for gender differences between mares, geldings and stallions across all components using the combined data set (i.e. all breeds). Where significant differences were identified by the Kruskal-Wallis, multiple comparison analysis was carried out.

Finally, Spearman rank order correlations of age against personality component scores were carried out on the whole data set, in order to test for the effect of age on personality.

4.2.5 Ethics and welfare

This experiment was assessed and approved by the Moulton College Research Committee. Within this experiment, there was no direct manipulation, contact or observation by the researcher with the study animals. All study horses were assessed by their owners/carers using a non-invasive assessment method which was therefore considered appropriate for use by laypersons. The husbandry and care of the horses were also under the control of the owners/carers and met the appropriate welfare standards.

Respondents were contacted passively either through the relevant breed society or through appropriate media (see Section 4.2.2) and were informed of the purpose of the study. All respondents had the choice of filling in the questionnaire and were informed that data were to be used as part of an aggregated data base and that their details would not be disclosed individually (see Appendix 8). Furthermore, individuals were given the option of whether

or not to include personal details (e.g. respondent's name, horse's name and age) when completing the questionnaire. Personal contact details were not required and were not requested from respondents. Respondents were also provided with contact details for the researcher and therefore had the opportunity to request further information on the project if desired.

4.3 Results

4.3.1 Differences across all eight breeds

The Kruskal-Wallis tests were carried out for each component by comparing the scores of all eight breeds (Table 4.2). The results indicated that there were significant differences in component scores between breeds for all six of the components ($\chi^2 > 29.10$; $P<0.001$)

4.3.2 Breed differences

A total of 28 different breed pair combinations were possible from the eight horse breeds. Variability within personality components was quantified by how many of the 28 pairs were significantly different. The highest levels of variability were found on *Excitability* (13/28 breed pairs significantly different) and *Anxiousness* (13/28 breed pairs significantly different). In contrast *Antagonism* (4/28 breed pairs significantly different) *Protection* (3/28 breed pairs significantly different) *Inquisitiveness* (3/28 breed pairs significantly different) and *Sociability* (6/28 breed pairs different) had much lower levels of variability. For each component the breeds were ranked according to their mean ranks calculated during the Kruskal-Wallis test. These are shown in Table 4.2, where the significant differences between breed pairs are also identified.

Table 4.2 provides a basic representation of each breed's typical personality. For example, on average the thoroughbred was rated highly on the components *Antagonism, Anxiousness, Activity* and *Sociability* and had moderate to low ratings on *Inquisitiveness* and *Protection*. In contrast the Irish draught horse had low scores on the components *Antagonism, Activity,* and *Inquisitiveness*, moderately low scores on *Anxiousness* and *Sociability* and a high score on *Protection*. Comparisons between these two breeds

identified that they were significantly different on all components except *Sociability* and *Protection.*

In contrast, interesting similarities were identified between the thoroughbred and the Arab which were only significantly different on *Protection* and *Antagonism*, as well as between the Arab and the Welsh ponies and cobs which only differed on *Sociability* and *Protection.* Furthermore, the Welsh ponies and cobs were not significantly different from thoroughbreds on any of the components.

Eight breed pair combinations did not show any significant differences on any of the components, these were; Shetlands versus Highlands, Appaloosas and American quarter horses; Irish draught horses versus Highlands, Appaloosas and American quarter horses; thoroughbreds and Welsh ponies and cobs; and finally, Appaloosas versus American quarter horses.

Comparisons are also displayed graphically in Figures 4.1 to 4.6, where the breeds are arranged in ascending order on each component. Additional graphical representation of the breed profiles are provided in Appendix 9.

Table 4.2: Average component scores for each breed with χ^2 values from Kruskal-Wallis test and summary of between breed differences.

	Median (IQR)								χ^2	Summary of between breed differences [y]
	IDH	**TB**	**Shet.**	**Arab**	**High.**	**WCP**	**App.**	**AQH**		
Antagonism	-1.26 (1.62)	-0.30 (2.13)	-1.15 (2.08)	-0.76 (2.03)	-0.88 (2.41)	-0.86 (1.59)	-0.89 (1.93)	-0.82 (2.15)	53.07***	TB^{a}, WPC^{ab}, AQH^{ab}, Arab.^{b}, High^{ab}, App.^{b}, Shet.^{b}, IDH^{b}
Anxiousness	2.93 (2.00)	3.96 (2.78)	2.80 (1.97)	3.59 (2.69)	2.70 (2.04)	3.67 (2.84)	2.67 (1.46)	2.72 (1.89)	121.22***	TB^{a}, Arab^{a}, WPC^{a}, IDH^{bc}, AQH^{bc}, Shet.^{c}, App.^{c}, High.^{c}
Activity	5.05 (2.15)	6.29 (2.30)	5.42 (2.48)	6.13 (2.25)	4.98 (2.31)	6.23 (2.09)	5.13 (2.44)	5.42 (2.43)	111.78***	TB^{a}, Arab^{a}, WPC^{ab} Shet.^{bc}, AQH^{bc}, App.^{c}, High.^{c}, IDH^{c}
Sociability	5.43 (2.40)	5.86 (1.98)	5.65 (1.99)	5.85 (2.11)	5.12 (1.85)	5.33 (1.86)	5.21 (2.29)	4.90 (2.85)	44.94***	Arab^{a}, TB^{ab}, Shet^{abc}, IDH^{bc}, WPC^{bc}, App^{bc}, High^{c}, AQH^{c}
Protection	6.77 (2.50)	6.53 (2.64)	6.76 (2.74)	7.22 (2.36)	6.82 (2.67)	6.35 (2.28)	6.69 (2.38)	6.44 (2.98)	29.10***	Arab^{a}, IDH^{ab}, Shet^{ab}, High^{ab}, App.^{ab}, TB^{b}, WPC^{b}, AQH^{b}
Inquisitiveness	7.83 (2.27)	-0.16 (3.31)	8.57 (2.61)	8.55 (2.46)	8.32 (2.61)	8.36 (2.21)	8.16 (2.46)	7.68 (3.00)	30.42***	Arab^{ab}, Shet^{ab}, High.^{ac}, TB^{ab}, Welsh.^{bc}, App.^{bc}, AQH^{bc}, IDH^{c}

[y] Medians significantly different (multiple comparisons test, $P<0.05$) where at least one superscript differs, breeds shown in rank order, highest to lowest (according to mean ranking in Kruskal-Wallis test, df = 7, *** $P<0.001$). IQR = Interquartile range; IDH = Irish draught horse; TB = thoroughbred; Shet. = Shetland pony; High. = Highland pony; WPC = Welsh ponies and cobs; App. = Appaloosa; AQH = American quarter horse

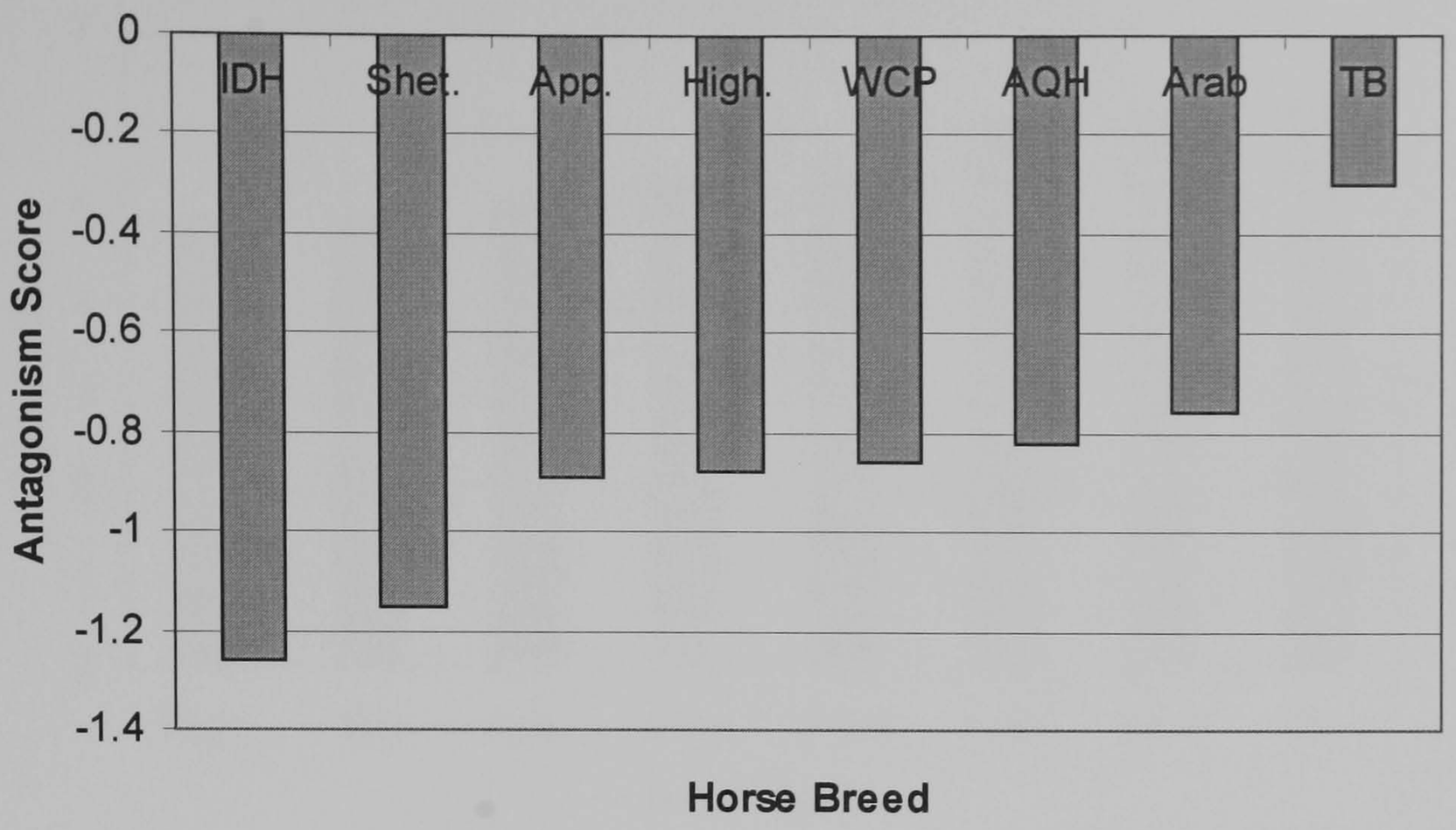


Figure 4.1 Median *Antagonism* scores for eight horse breeds assessed using the Horse Personality Questionnaire.

IDH = Irish draught horse; TB = thoroughbred; Shet. = Shetland pony; High. = Highland pony; WPC = Welsh ponies and cobs; App. = Appaloosa; AQH = American quarter horse.

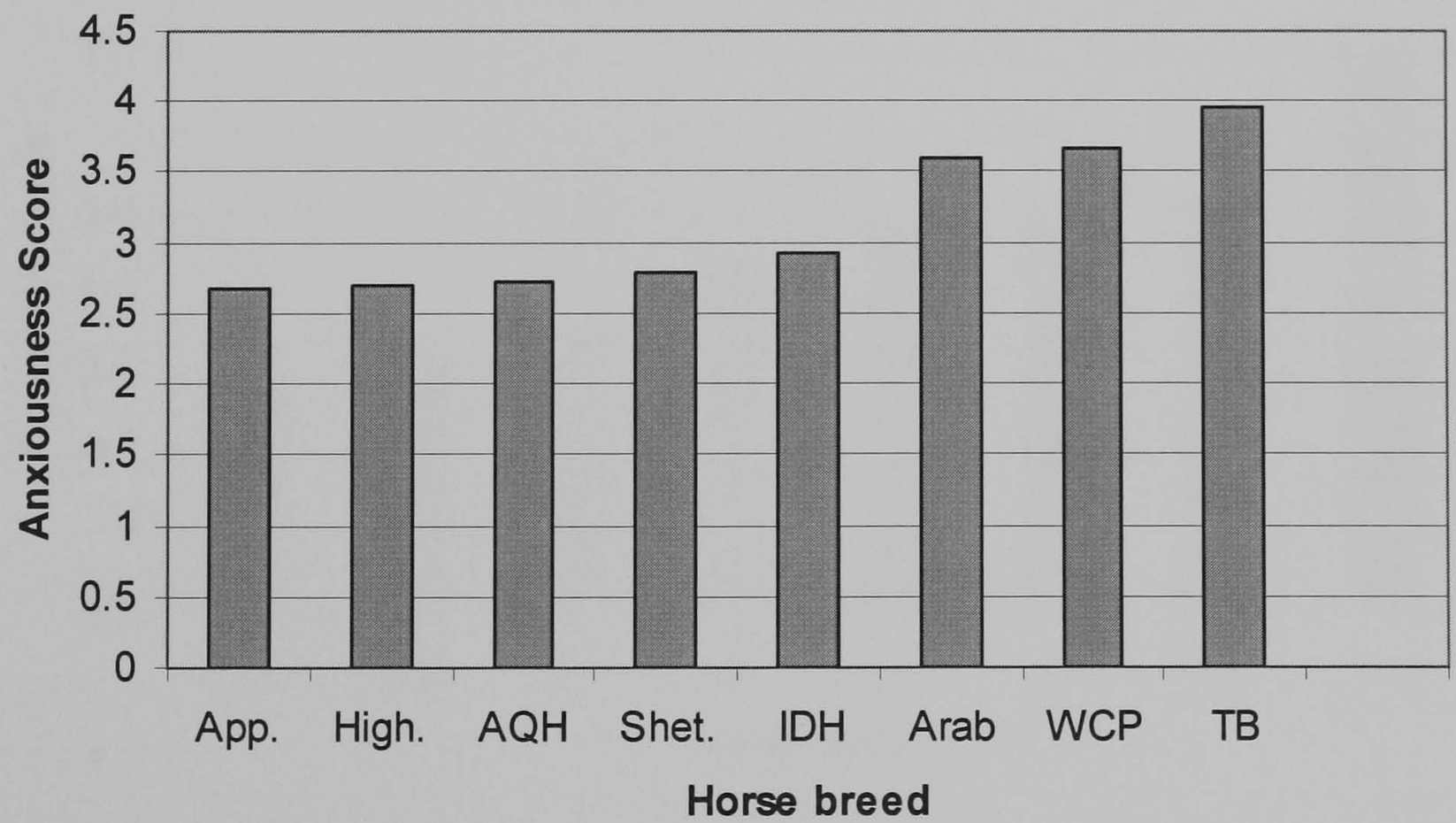


Figure 4.2 Median *Anxiousness* scores for eight horse breeds assessed using the Horse Personality Questionnaire.

IDH = Irish draught horse; TB = thoroughbred; Shet. = Shetland pony; High. = Highland pony; WPC = Welsh ponies and cobs; App. = Appaloosa; AQH = American quarter horse.

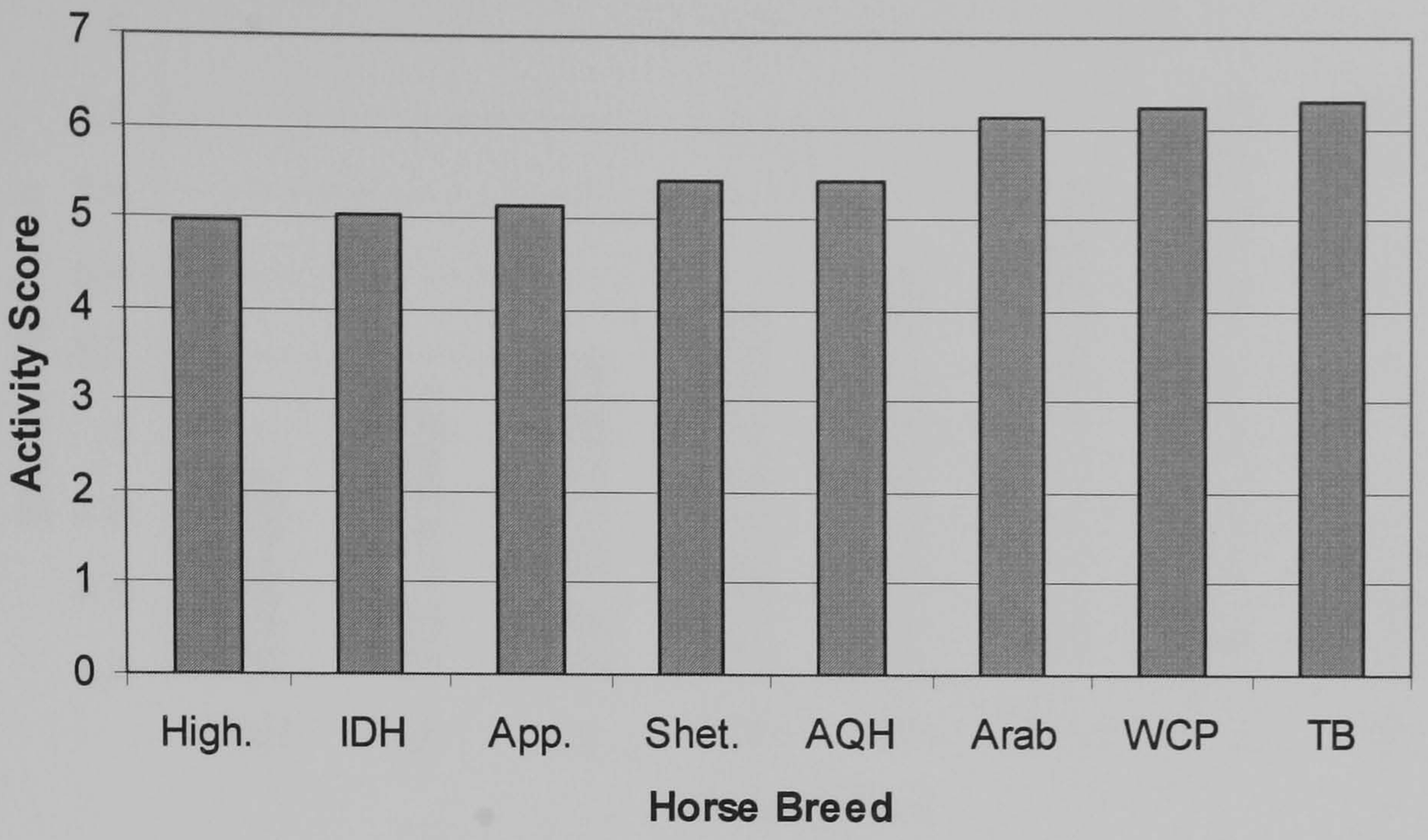


Figure 4.3 Median *Activity* scores for eight horse breeds assessed using the Horse Personality Questionnaire.

IDH = Irish draught horse; TB = thoroughbred; Shet. = Shetland pony; High. = Highland pony; WPC = Welsh ponies and cobs; App. = Appaloosa; AQH = American quarter horse.

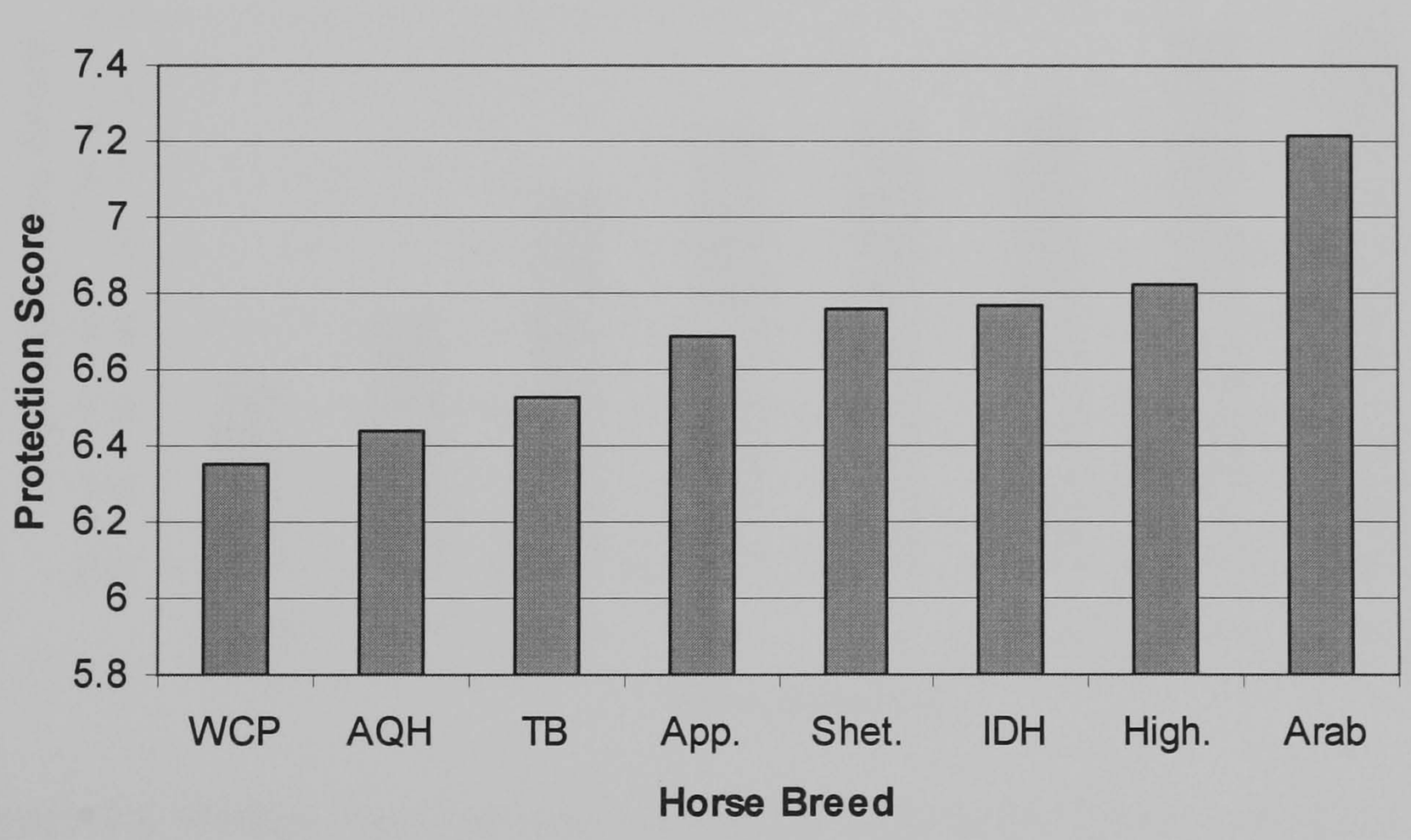


Figure 4.4 Median *Protection* scores for eight horse breeds assessed using the Horse Personality Questionnaire.

IDH = Irish draught horse; TB = thoroughbred; Shet. = Shetland pony; High. = Highland pony; WPC = Welsh ponies and cobs; App. = Appaloosa; AQH = American quarter horse.

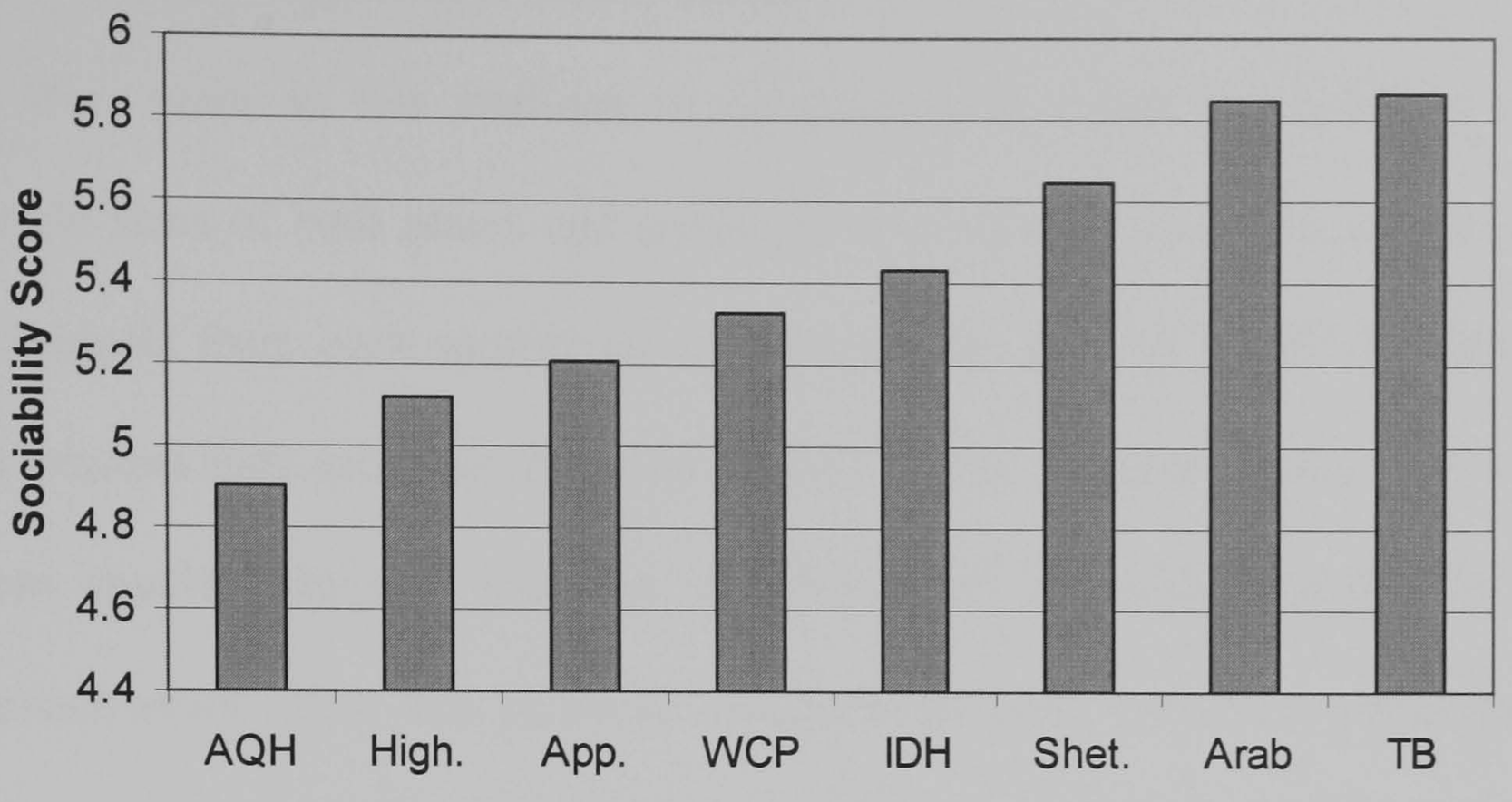


Figure 4.5 Median *Sociability* scores for eight horse breeds assessed using the Horse Personality Questionnaire**.**

IDH = Irish draught horse; TB = thoroughbred; Shet. = Shetland pony; High. = Highland pony; WPC = Welsh ponies and cobs; App. = Appaloosa; AQH = American quarter horse.

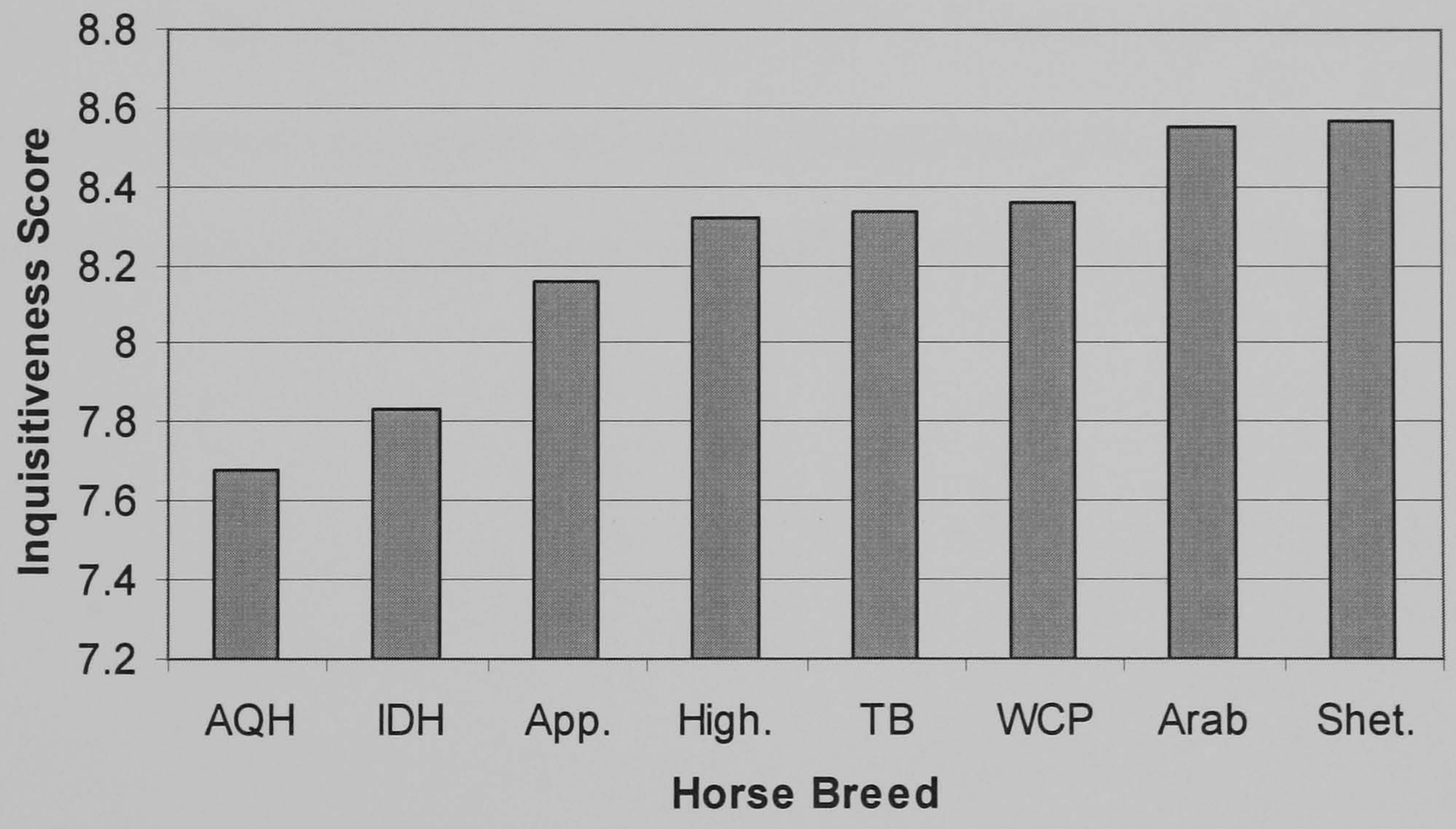


Figure 4.6 Median *Inquisitiveness* scores for eight horse breeds assessed using the Horse Personality Questionnaire.

IDH = Irish draught horse; TB = thoroughbred; Shet. = Shetland pony; High. = Highland pony; WPC = Welsh ponies and cobs; App. = Appaloosa; AQH = American quarter horse.

4.3.3 Gender differences within breeds

As there were so few stallions in comparison to mares and geldings (40:561:618) the sample sizes of both mares and geldings were reduced by randomly selecting 10% of the individuals from each sample group reducing the ratio to 40:55:67. This was done using the random data selection function on SPSS. The Kruskal-Wallis test was performed on these smaller samples and was significant ($\chi^2 > 10.30$, $P<0.01$) on *Sociability* and *Inquisitiveness*. Post hoc multiple comparisons tests were then performed on data from these components (six comparisons per component) and identified a significant difference between males and females. No significant difference was identified between geldings and stallions. These results are summarised in Table 4.3.

4.3.4 The association between age and personality score

Very weak, but significant ($r_s <\pm 0.20$, $P<0.05$) Spearman rank order correlations were identified between personality and age on all components except *Anxiousness.* Correlations were all negative except on *Protection* which appeared to increase with age see Table 4.3.

Table 4.3: Age-personality correlation coefficients (r_s) and gender differences tested using Kruskal-Wallis test and multiple comparisons (mare and gelding sample sizes reduced [x])

Component	Correlation	Gender differences[x]	
	Age (r_s)	Kruskal-Wallis (χ^2)	Multiple comparisons[y]
Antagonism	-0.09**	1.47	N/A
Anxiousness	-0.06	3.11	N/A
Activity	-0.06*	4.24	N/A
Sociability	-0.12***	10.31**	Stallion[a], Gelding[a], Mare[b]
Protection	0.09**	0.34	N/A
Inquisitiveness	-0.18***	12.38**	Stallion[a], Gelding[a], Mare[b]

*$P<0.05$;**$P< 0.01$;***$P<0.001$

[x] stallion n = 40, gelding n = 67 and mare n = 55.

[y] Significantly different (multiple comparisons test, $P<0.05$) where at least one superscript differs. Genders shown in rank order, highest to lowest (according to mean ranking in Kruskal-Wallis test, df = 2).

4.4 Discussion

The results demonstrated that differences in personality exist between the eight horse breeds assessed in this study. Differences between breeds varied across the six personality components, with *Anxiousness* and *Activity* demonstrating the most variability and *Antagonism, Inquisitiveness* and *Protection* showing the least amount of variability. Gender differences were also identified in *Sociability* and *Inquisitiveness*. The association between age and personality score was significant on all components except *Anxiousness*. Correlation coefficients were small; inferring that the relationship between age and personality score was not very strong. Nonetheless significant breed differences in personality were still identified, despite these within breed variations. The following discussion explores between breed variation and more general age and gender effects on horse personality on each component. The results are discussed in terms of the potential influence of selective processes on breed personality.

4.4.1 Anxiousness and Excitability

The components *Anxiousness* and *Activity* showed the highest level of variation between breeds. The thoroughbreds, Arabs, and Welsh ponies and cobs were ranked as the top three breeds and were not significantly different from each other on these two components. These results infer that these breeds are behaving similarly in terms of *Activity* and *Anxiousness* and that they are more active and anxious than the other breeds included in this study. Similarly, Hausberger *et al.* (2004) found that thoroughbreds and Arabs were the most reactive horses of a sample of 16 horse breeds when tested using a bridge test. The high levels of *Anxiousness* and *Activity* demonstrated by these breeds could partly be attributed to the Arab ancestry of both the thoroughbred (Bowling & Ruvinsky, 2000) and Welsh ponies and cobs (Draper, 2001; Foster, 2005), which will now be discussed.

The thoroughbred breed was founded during the 1700s using Arab stallions which were bred with English native mares, (Bowling & Ruvinsky, 2000; Draper, 2001; Foster, 2005). Similarly the Welsh mountain pony (Welsh Section A) has been influenced by the addition of Arab and eastern bloodlines, introduced by the Romans (Draper, 2001; Foster, 2005) and more recently (within the last 200-300 years) through Arab stallions reportedly being allowed to run with free-ranging herds of Welsh mountain ponies (Draper, 2001; Foster, 2005). The Welsh ponies and cobs are all descended from the Welsh mountain pony. Furthermore, the Welsh section B was founded by the crossing of Welsh mountain pony mares with thoroughbred and Arab stallions (Draper, 2001; Foster, 2005). Thus these three breeds have a linked ancestry and it seems that this may have resulted in breeds that are still very comparable in their levels of *Activity* and *Anxiousness*.

Although historical records indicate towards the linked ancestry of these breeds evidence of continued genetic relatedness and similarity is still limited. One example of such a study is provided by Bowling and Ruvinsky (2000) who explored the relatedness of ten horse breeds (including the Arab and thoroughbred) and Przewalski's horse (*Equus przewalskii*). Genetic distances between the ten breeds and Przewalski's horse were estimated, based on 38 loci. Not surprisingly the Przewalski's horse was the most dissimilar in any of the paired comparisons, with a value of at least 0.308. In contrast the thoroughbred and Teke were identified as being the most similar (0.041 ± 0.01). Such a similarity was not unexpected due to the recent use of thoroughbred stock in Teke breeding programmes. The genetic distance between thoroughbreds and Arabs was also relatively small (0.105 ± 0.02). A subsequent dendrogram of these data identified thoroughbreds, Arabs and Tekes as having the closest relationships with several of the other breeds (e.g. Lipizzaner, Morgan and Iberian) branching off from this set. Similar data relating to the Welsh breeds

and their genetic relationships with the Arab and thoroughbred do not yet appear to be available.

Genetic factors have been noted by Hausberger *et al.* (2004) to have a great influence on the neophobic responses of horses. These are most likely to be associated with *Anxiousness* and *Activity*. Furthermore, sire influences have been found to affect horse emotionality (Wolff *et al.*, 1997; Houpt & Kusunose, 2000) and the tendency to develop stereotypic behaviour (Houpt & Kusunose, 2000). Due to the proposed links with stress (Mills *et al.* 2002) and therefore anxiety, it is possible that stereotypic behaviours and emotionality are influenced by the components *Anxiousness* and *Excitability*.

The high level of variability in *Anxiousness* and *Excitability*, between breeds suggests that the impact of artificial selection has been great on these particular characteristics. As such this indicates that artificial selection has developed breeds that show a wide range of *Activity* and *Anxiousness*, creating breeds that are behaviourally tailored for their function. For example, McGreevy and Thompson (2006) noted that, for the purposes of racing, thoroughbreds need to be highly reactive to stimuli and therefore have heightened flight responses. As a result, they are quick off the starting line. Similarly, it seems likely that draught and multipurpose breeds, such as the Irish draught horse and the Highland pony, would be selected for low levels of *Activity* and *Anxiousness*. The overall greater strength and bulk of such breeds could otherwise make them difficult and dangerous to handle. Therefore selection of appropriate personality types may have aided the management and handling of horses and has made them more suitable for their relevant functions.

4.4.2 Sociability and Protection

The lower variability between breeds on these components may indicate that the process of artificial selection has influenced these characteristics less than those of *Anxiousness* and *Activity*. *Sociability* may be of importance in terms of the ease of housing horses together, high sociability would therefore be beneficial in all breeds. Furthermore, high levels of sociability would be expected due to the naturally social nature of the species (Clutton-Brock, 1999). It would, therefore be less likely to have been specifically selected for than *Anxiousness* or *Activity* as there would be less need for such variation.

Protection also demonstrated low levels of variance between breeds. The average scores for all breeds were high indicating that all breeds showed high levels of protection. The breeds found to be most protective were the Arab and the Irish draught horse. The least protective were the thoroughbred, American quarter horse and the Welsh ponies and cobs.

It is interesting to note that the rank order of breeds on *Protection* is noticeably different when compared to the rankings on the other components. The thoroughbred, Arab and Welsh ponies and cobs are regularly seen clumped together on the other five components, and although not statistically similar in all cases, they group at the same ends of the ranking scales. In contrast, on *Protection* the Arab broke away from this group and ranked highest, whereas the thoroughbred and the Welsh ponies and cobs were ranked in the bottom three. Given the reported ancestry of these breeds (see earlier) and their apparent similarity on the other components, it seems unusual that that they should become segregated on *Protection*. This could be suggestive of a maternal effect and it would be interesting to investigate why this division has occurred. The majority of Arab influence on both the thoroughbred and the Welsh ponies and cobs has reportedly been as a result of stallions being put to native mares, and this sire influence may have affected the scores for

the other five components. As *Protection* has a mothering and nurturing aspect, perhaps the maternal influence (genetic and/or environmental) is greater than that of the sire. Maternal influence has previously been found to affect some behaviour (McAdam *et al.*, 2002). For example strong maternal influences were found in dominant-aggressive behaviour in the English cocker spaniel (Pérez-Guisado *et al.*, 2006) and cross-fostering studies on Rhesus macaques identified maternal influences on reactive behaviours (Suomi, 1987). Further research should look towards exploring the maternal influences, both genetic and environmental, on this component.

4.4.3 Antagonism and Inquisitiveness

The average scores for *Antagonism* for all breeds were negative, indicating that in general individuals were receiving high scores on the more desirable traits (reliable, subordinate and equable). The variation between breeds on this component was low, with the most significant differences occurring between those breeds at opposite ends of the ranking scale. The thoroughbred and Welsh ponies and cobs were ranked highest and the Shetland pony and Irish draught horse were ranked as the lowest on this component.

In Chapter 3 *Antagonism* was identified as being comparable to the component *Dominance* which has been identified in other animal species (e.g. Chimapnzees, King and Figueredo, 1997). Research on *Dominance* in dogs (Strandberg *et al.*, 2005) and chimpanzees (Weiss *et al.*, 2000; Weiss *et al.*, 2002) has identified that *Dominance* is heritable to some extent in these species. It is therefore possible that low levels of *Antagonism* may have been selected for and would be desirable in all horse breeds. Reduced levels of dominance may aid the management of the species by creating more handleable animals. The selection of low levels of *Dominance* forms part of the general domestication process. For example, Fraser and Broom (1997) noted the importance of the human caretaker being dominant in those

species that have a dominant-subordinate type of social structure, especially when the animals can be potentially dangerous as adults, as can be seen in horses.

Low levels of variability were also identified in *Inquisitiveness* with only three between breed differences identified. These indicated that the Irish draught horse was significantly less inquisitive than the Shetland pony, Arab and thoroughbred, but that the remaining breeds were more alike.

The opportunistic element of *Inquisitiveness* may in some cases be disadvantageous, in that highly inquisitive horses may be more likely to escape, resulting in a general trend towards selection for low *Inquisitiveness*. Horses with higher levels of *Inquisitiveness* may, however, be selected for cross-country trials where a quick-witted and opportunistic horse would be better able to cope with the complicated elements of the course. The thoroughbred was ranked fourth highest on *Inquisitive* and this breed and its crosses are regularly used for this event (Bowling & Ruvinsky, 2000; Draper, 2001). Further research could explore the personality of performance horses such as these, to investigate the personality types of successful versus unsuccessful horses.

4.4.4 Breed typical personalities

Although data were limited to eight horse breeds, those that were used were representative of some of the major types (i.e. Lightweight, draught and pony). It was noted that breeds from within the same categories tended to have more similar personality types, thus these results may allow for some generalisability of personality type to other breeds within these categories. In addition, the results supported the beliefs held by many horse enthusiasts that horse breeds differ in their typical personalities. Furthermore, many of the traditional views

of the breeds assessed have also been supported, despite some variation in personality associated with gender and age.

Hayes (1998) conducted a survey of 50 veterinarians and trainers, asking for comments on ten horse breeds and six characteristics. These were trainability, work-ethic, temperament, 'when asked to do something the horse does not want to do', response to pain, fearfulness and flight. Breeds that were included in both the present study and Hayes (1998) were the Appaloosa, Arab, American quarter horse and thoroughbred. The descriptions of all four of these breeds by Hayes match very well to the results of the present study. For example the Appaloosa was found by Hayes to be easy going and low on both fearfulness and flight. The present study found the Appaloosa to have moderate to low scores on all of the components in comparison to the other breeds and thus supports Hayes' description. Additionally the Arab was described as a fast learner, energetic, playful and reactive for both fearfulness and flight, this is comparable to the results of the present study, where the Arab scored highly on five components and moderately high on *Antagonism* in comparison to the other breeds.

The results also compared well to the society descriptions of the breeds. For example, the Irish draught horse was described as an intelligent breed with a gentle nature, docility and sense (Irish Draught Horse Society, 2006), and the Highland pony was described as having a '*kindly nature and even temperament*' (Highland Pony Society, 2006) such that the breed was not considered to be flighty or highly reactive. Both of these breeds, when compared to the other breeds, had comparatively low scores on all of the components, except *Protection* where they were ranked highly, thus providing support for these claims.

4.4.5 Gender differences within breeds

Although gender differences were identified, they were restricted to the components *Sociability* and *Inquisitiveness*. On both of these components males scored more highly than the females. Significant differences between stallions and geldings were not, however, identified, suggesting that castration has no or a limited effect on these components.

Gender differences in personality have previously been identified in other species. These have included chimpanzees (Buirski *et al.*, 1978), vervet monkeys (McGuire *et al.*, 1994), spotted hyenas (Gosling, 1998) and humans (Budaev, 1999; Costa *et al.*, 2001). It is therefore not unexpected to find some level of variance in horses. Wolff *et al.* (1997) and Hausberger *et al.* (2004), however, did not identify gender differences in the reactions of horses during behavioural tests designed to assess individual differences in emotionality and learning (e.g. bridge test and novel object test). Behaviour tests, as used by Wolff *et al.* (1997) and Hausberger *et al.* (2004), are most likely to map onto the personality components of *Anxiousness* and *Activity*. These components showed no gender differences and do not, therefore, directly contradict Wolff *et al.* (1997) and Hausberger *et al.* (2004). The components that demonstrated gender differences, *Sociability* and *Inquisitiveness,* contain traits that may be more difficult to assess during conventional behavioural tests (i.e. playful, popular, social, inquisitiveness and curiosity) and therefore may not have been directly assessed by Wolff *et al.* (1997) or Hausberger *et al.* (2004).

Males were found to score more highly than females on *Sociability* and *Inquisitiveness*. *Sociability* comprised positive loadings of the traits popular, sociable and playful, thus inferring that males spent longer playing and a greater amount of time with other individuals (this may be other horses or humans). *Inquisitive* is positively associated with the traits curious and opportunistic, therefore inferring that males are more curious and

more capable of noticing and/or seizing opportunities. These results are supported by Feh (1988) who noted that play behaviour accounted for 23.9% of the variance in social behaviour exhibited by a bachelor herd of semi-free-ranging Przewalski horses. Furthermore McDonell and Poulin (2002) noted that bachelor males and harem stallions were more often seen exhibiting play behaviour than mares, indicating that play has an important role in the behaviour of free-ranging horses, but even more so in males.

Sexual dimorphism in personality has been suggested in great tits as a result of fluctuating selection (Dingemanse *et al.,* 2004). Such that in a wild population of great tits personalities appeared to fluctuate markedly between years and that these fluctuations were different between males and females. For example, in the first and last study seasons (of three) slow females and fast males survived better, whereas in the middle season, the converse was observed. In this case, such fluctuation was suggested to have been linked to variation in resource availability, for example, with the mass-seed crops (masting) of local beech (*Fagus sylvaticus*) populations which were an important winter food source for the birds. The first and last seasons had followed poor masting winters, whereas the middle season followed a winter where the beech had masted and many birds survived. Due to males being socially dominant over females, over-winter access to food was more likely to determine female survival, whereas ability to defend a territory was more important to male viability. These results indicated that fast males survive better when breeding populations are high, while fast females persist when food is limiting in the winter.

Such selection processes acting on personality and the resulting fitness payoffs have recently been discussed by Dall (2004), Dall *et al.* (2004) and Dingemanse and Réale (2005). As a result, previous assumptions that personality and individual differences in behaviour were "*non-adaptive variation surrounding (possible) adaptive population-*

average behaviour" (Dall *et al.*, 2004, p. 734) is slowly being replaced with a more behavioural ecologically orientated view that incorporates Darwinian theories. This developing view of personality proposes that personality differences can be selected for and may require the use of game theories to further understand the evolution and maintenance of personality (Dall *et al.*, 2004).

From an evolutionary perspective, male horses may have needed to be more inquisitive and social than females. Wild and feral horses often form bachelor groups of mature and sub-adult males, where play behaviour has an important role (Feh, 1988; McDonnell & Haviland, 1995). It is likely that in such a situation, play enables the young males to develop their fighting skills as well as assess the strength of their future opponents, in preparation for contests over females. It is feasible that such a social system could have applied selection pressure on the more opportunistic and curious males. Such individuals may be more able to seize opportunities when they arise, such as the potential to recruit juvenile females or take over a herd from an ailing stallion. As a result it may be that higher levels of sociability and play in males may also be linked to breeding success. Those males that form allegiances with other males in the group and spend time play fighting would perhaps have better access to resources, fewer aggressive interactions, and would have a greater knowledge of the abilities of their future opponents as well as having a greater opportunity to improve their own skills.

Gender differences on *Antagonism* would perhaps be expected in a gregarious species where the males are often the most dominant individuals in the herd (Arnold & Grassia, 1952; Clutton-Brock, 1999) or in the case of bachelor herds, when they are regularly re-assessing dominance status (McDonnell and Haviland, 1995). Arnold and Grassia (1952) observed the social behaviour of thoroughbreds and found that the stallions tended to

vocalise, lead and mix with other individuals more than the mares they were grouped with. They were also the most dominant in terms of aggressive interactions. As general horse behavioural ecology would appear to predict a significant difference in *Antagonism* between genders, it seems unusual that this was not identified by this study.

The gender differences identified on *Sociability* and *Inquisitiveness* were only between the two male categories and females, no significant differences were identified between geldings and stallions. This indicated that castration does not have an effect on these components. Similarly Wolff and Hausberger (1996) identified that there was no significant difference between learning abilities between stallions and geldings. Castration of young male horses is generally thought to have a significant effect on horse behaviour resulting in animals that are considered more manageable than intact males in terms of behaviour and practicality. McGreevy (2004) and McDonnell (2005) do however discuss reports of retained stallion-like behaviour in geldings. For example, McDonnell (2005) highlights the study by Line *et al*. (1985) who reported that of 140 geldings that had been castrated three to nine years earlier, 20-30% were still reported to perform stallion like behaviours towards other horses and 5% continued to perform stallion-like aggressive behaviour towards handlers. Furthermore, where castration had been carried out specifically to cease objectionable or unmanageable stallion-type behaviour, owners reported that castration had effectively reduced stallion like aggressive behaviour towards people in 60-70% of cases and towards horses in 40% of cases. Such evidence indicates towards behavioural differences between geldings and stallions yet this does not appear to be the case with reference to personality. Further research may benefit from a more detailed focus on the effects of castration on horse personality.

In general the gender differences exhibited in these breeds were few. Those differences that were identified may be as a result of evolutionary processes that have selected for different attributes in males and females. Such suggestions are at this stage highly speculative, but as has been identified by a range of authors (e.g. Bouchard & Loehlin, 2001; Dall, 2004; Dingemanse, & Réale, 2005; McElreath & Strimling, 2006) the evolutionary selection pressures on personality provide an interesting and rapidly growing area of research.

4.4.6 The association between age and personality score

A significant, but weak, relationship between age and personality score was shown on all components except *Anxiousness*. With the exception of *Protection*, all significant correlations were negative. This inferred that personality scores were slightly lower for older horses, but that older horses were rated more highly on *Protection*. The general decrease in personality scores suggests that horses become more affable and less active as they get older, although these changes appear to be quite small.

With respect to general personality theory, there are two main schools of thought with regards to age-personality relationships. According to the five-factor theory, personality traits develop throughout childhood and become stable during early adulthood, though personality could change again in old age due to cognitive decline (McCrae & Costa, 1999). In contrast the contextual perspective on personality and change predict that personality changes throughout adulthood due to exposure to different environmental influences (Srivastava *et al.*, 2003). The definition of personality provided by Pervin and John (1997) however, places emphasis on the consistency of behaviour as an integral part of personality. Behavioural consistency within years has previously been identified in horses by Visser *et al.* (2001) who repeatedly assessed 41 horses during the first 24 months

of life. Consistency between years at this young age was not, however, demonstrated suggesting that personality may be more unstable during early ontogeny becoming more stable in adult hood. It is therefore predicted that any variation in an individual's personality over time should be minimal, as has been demonstrated by these results.

The strongest association between personality and age was identified on the components *Sociability* and *Inquisitiveness*, with scores showing a slight decline with age. It is possible that as horses mature, the frequency of social behaviours, such as play, may decrease and therefore result in decreased scores of *Sociability*. Levels of activity are also likely to decrease with age (Boguszewski & Zagrodzka, 2002), due to changes in physical fitness, which may also result in a decrease in play behaviours. Furthermore, life experiences may aid a horse's understanding of the world. The occurrence of inquisitive behaviours may be reduced as novelty in the environment reduces with time and experience. The apparent decrease in *Inquisitiveness* is supported by the findings of Mader and Price (1980) who identified a significant negative correlation between horse age and rate of learning in a discrimination task. In addition this can be linked to the decrease in *Activity* in older horses due to the link with the trait intelligence on this component. It should however be noted that the relationships identified in this study are weak and in most cases did not account for a large amount of the variability.

There is mixed evidence for the effect of age on personality within the literature. Human studies appear to show both plasticity (McCrae, 2002) and change (Srivastava *et al.*, 2003) during adulthood. Within the animal literature age has been shown to affect; *Playful/Curious* in vervet monkeys (McGuire *et al.*, 1994); *Anxiety* in rats (Boguszewski & Zagrodzka, 2002); learning in horses (Mader & Price, 1980); agonistic activity and grooming behaviours in rhesus macaques (Suomi *et al.*, 1996) and *Extraversion* in gorillas

(Kuhar *et al.*, 2006). Such relationships are, however, uncommon. For example, Kuhar *et al.* (2006) tested for associations between age and four personality factors in gorillas. They identified that *Extraversion* was significantly negatively correlated with age, but no other significant relationships were identified. Age-personality relationships are therefore not consistent across personality components.

It should, however, be noted that the age effects identified in this study are exploratory and are not conclusive. The data were not gathered as part of a longitudinal study that followed individuals through their development and should be interpreted with care. A detailed longitudinal study of horse personality through different developmental stages would provide more reliable and conclusive results.

4.4.7 Generalisability of data

Due to the large sample sizes obtained for each breed these data were evaluated as being representative for the selected breeds. Although environment and individual conditions will have varied greatly between the individuals, this will have made the sample more representative of the general population. The conclusions drawn from these data are, therefore, limited in terms of generalisation to other horse breeds. The breeds used in the study were, however, selected to provide a representative sample of the overall horse population. For example, draught, pony, warm (e.g. Appaloosa) and hot (e.g. thoroughbred and Arab) blood breeds were included. The clumping together of similar breeds in terms of personality type, as discussed in earlier sections, indicated that breeds of similar function have similar personality types. It would therefore be interesting to explore further breed personality types to assess whether other breeds would clump into their appropriate categories. Furthermore, additional studies could begin to explore similarities in personality between horses from the same genetic lines, and further explore both the

paternal and maternal effects on personality. This could link to the work of Hausberger *et al.* (2004) who have already indicated that there may be some paternal control over personality in horses. The use of the HPQ in such a study, may allow for large numbers of individuals to be involved as the HPQ is a relatively simple and quick method of assessment that can be completed by a large volume of people, as was demonstrated by this study.

The placement of the HPQ onto the internet as part of its distribution did however bring potential complications in terms of data generalisation. The placement of the HPQ on the internet and international advertisement through an online equine magazine meant that HPQ was opened up to international respondents. Details regarding place of origin were not, however, requested as part of the HPQ, it was therefore impossible to identify where the online responses had originated from. Had this been possible, it would have been interesting to explore any personality differences between sub-populations of breeds (e.g. English thoroughbred versus American thoroughbred) as different trends in selection may have resulted in slight variations in personality type. The use of the HPQ as a widespread method of assessment may therefore be of great use in further areas of horse personality research.

4.5 Summary

The results of this study have provided further evidence for breed typical personalities and are supported by findings of other studies (Hausberger & Muller, 2002; Hausberger *et al.*, 2004). General Aim IV (Section 1.1) has therefore been met. These results also demonstrated that horse personality can be linked to real-world outcomes (Criterion Two, Gosling and Vazire 2002) (i.e. predicted breed differences) and provides further evidence for the efficacy of the HPQ.

Moreover, the breed differences and the similarity of personality profile between those breeds with a linked pedigree, provides further support for the heritability of behaviour. This in turn raises questions about the maternal and paternal effects on inheritance of behaviour and personality in horses. It is hoped that these results will spark further and more detailed research on the heritability of horse personality and behaviour.

The findings of this study and others that explore behavioural differences between breeds should be made available to horse owners and handlers. Information regarding a breed's typical behaviour and personality may allow for more informed decisions during the selection of horses for both leisure riding and other areas of equitation. This may be of little use if personality can not be used to predict behaviour. It is therefore important that the potential to predict behaviour from personality ratings is tested. This will now be explored in Chapter 5 in order to further validate the HPQ.

5 Use of Personality Scores to Predict Behaviour in Horses

5.1 Introduction

As discussed in Chapter 2, personality, by definition, encompasses consistency in behaviour and the way an individual feels and thinks (Pervin & John, 1997). Therefore knowledge of an individual's personality should allow for predictions of how that individual would behave in specific situations, as their behaviour should show consistency across time and situations (Epstein, 1979; 1983). Gosling and Vazire (2002), in their review of animal personality research, emphasised the importance of being able to predict behaviour from personality assessments stating that in order to demonstrate reliability "*assessments must predict behaviours and real-world outcomes*" (Criterion Two, Gosling & Vazire, 2002, p. 608). Chapters 3 and 4 demonstrated both the reliability and validity of the HPQ, yet Criterion Two has yet to be fully met. This chapter explores the use of the HPQ in the prediction of behaviour in individual horses.

The ability to predict future behaviour in working animals has great potential in the selection of suitable individuals for specific disciplines as well as in the prediction of future performance (Mills, 1998; Visser *et al.*, 2003a). This may be particularly useful in the equine industry as horses are used in a large variety of sports as well as in the police and defence forces.

Although interest in horse personality has increased and the potential applications have been discussed (Mills, 1998), few studies have directly looked at personality-behaviour associations and predictions in horses. Some studies have, however identified consistency

in behaviour (Visser *et al.*, 2001; McCall *et al.*, 2006) and, as has been identified in Chapter 3, links with rated personality and behaviour recordings (Momozawa *et al.*, 2003). Nevertheless few studies have directly looked at using personality as a predictor of behaviour or performance. One study by Visser *et al.* (2003a) has, however, attempted to predict show-jumping performance from the behaviour of young horses. Using behavioural tests the following were quantified in horses during the first two years of life; emotionality, reactivity to humans and learning ability. No significant correlations were found between the behavioural measures used in the personality tests and the show-jumping performance of the horses at three years of age. The authors indicated that the personality characteristics assessed using the behavioural tests were not as relevant to show jumping performance as had been predicted and suggested further development of more suitable behavioural tests.

The study by Visser *et al.* (2003a) demonstrated the difficulty in assessing personality using behavioural tests. Although they can be reliable and show consistency they are restrictive in the range of personality characteristics that they can assess. In contrast, the trait rating method provides an opportunity to assess the animal as a whole rather than focusing on specific characteristics. The HPQ is an example of such a rating method and has been shown to be both reliable and valid (as demonstrated in Chapters 3 and 4). The HPQ would therefore be a suitable tool for the assessment of personality for predicting future behaviour in a variety of contexts, for example, behaviour and learning tests.

The aim of this experiment was to use the HPQ to assess the personality of individual horses and to predict subsequent behaviour in three behavioural tests (General Aim IV, Section 1.1) and therefore meet Gosling and Vazire's (2002) Criterion Two. The behaviour tests used were an arena test, a learning test and a release test. These tests were chosen as they were allowed for the expression of a range of personality characteristics. Both the

arena and the learning tests were adapted from Le Scolan *et al.* (1997) and Wolff *et al.* (1997). The arena tests observed the behaviour of individual horses when released into a familiar but enclosed arena and were isolated from their conspecifics. This aimed to reflect the *Sociability* of the individual. Using a simple instrumental task, the horses' ability to learn was assessed by measuring the time taken to complete a simple task. *Inquisitiveness* is defined by the traits curious and opportunistic and the component *Activity* is partly defined by the trait intelligent (see Table 3.4 and Table 3.5). These components were therefore predicted to be associated with the behaviours measured in the learning test. The third test observed horses during handling and post release to a field containing familiar conspecifics. Partly defined by the traits activity and excitable, *Activity* was predicted to be associated with ease of handling. It was also predicted that the occurrence and type of reactions exhibited by the horses upon returning to their conspecifics would be correlated with *Sociability*.

The behaviour tests were chosen as they were simple, practical and quick to perform thus allowing for their potential application in other equine establishments. Both the arena and the learning test have demonstrated individual differences in horse behaviour (Le Scolan *et al.*, 1997; Wolff *et al.*, 1997), and were evaluated as being suitable for testing predictions of personality-behaviour correlations. Arena and open-field tests have been used in a number of studies to assess the reaction of individuals to isolation from conspecifics (Kilgour, 1975; Boissy, 1995; Le Scolan *et al.*, 1997; Wolff *et al.*, 1997).

The following predictions (summarised in Table 5.1) are consistent with the implied meaning and structure of the personality components, as described in Chapter 3 (see Tables 3.4 and 3.5), and describe the general personality-behaviour predictions of this study.

1) *Sociability*, which is defined by traits such as playful and popular, was predicted to be associated with those behaviours exhibited when isolated from conspecifics as well as when returned to a familiar social group. Behaviours may include high levels of activity, vigilance and vocalisation and a low occurrence of standing.
2) *Activity*, defined by the traits active, excitable and intelligent was predicted to be associated with ease of handling, time taken to learn a simple task and the behavioural reaction to a potentially stressful situation (e.g. isolation from conspecifics).
3) *Inquisitive*, which is defined by the traits curious and opportunistic was predicted to be associated with learning ability (time to complete a simple instrumental task) as well as exploration behaviours.
4) *Anxiousness*, defined by traits such as tense, fearful and insecure, was predicted to be associated with behaviours exhibited during social isolation. The components *Protection* and *Antagonism* were not assessed for behaviour correlates for the purpose of this study.

Table 5.1: Predicted associations of personality component scores with behaviours recorded during three behavioural tests

Test	Behaviour	Personality Components			
		Anxiousness	***Activity***	***Sociability***	***Inquisitiveness***
Arena	Gate	+	*n/a*	+	*n/a*
	Activity	+	+	+	*n/a*
	Exploration	*n/a*	*n/a*	*n/a*	+
	Vigilance	+	*n/a*	+	*n/a*
	Vocalisation	+	*n/a*	+	*n/a*
	Spook	+	+	*n/a*	*n/a*
	Stand	-	-	-	*n/a*
Learning	LT	*n/a*	*n/a*	*n/a*	-
	LTO	*n/a*	-	*n/a*	-
	Total T	*n/a*	*n/a*	*n/a*	+
Release	EoH	*n/a*	-	*n/a*	*n/a*
	LTA	-	*n/a*	-	*n/a*
	Affiliative	*n/a*	*n/a*	+	*n/a*
	Antagonistic	*n/a*	*n/a*	-	*n/a*

LT, latency to touch; LTO, latency to open; Total T, time touching box; EoH, ease of handling; LTA, latency to approach; Affiliative, number of affiliative interactions; Antagonistic, number of antagonistic interactions, *n/a* = not applicable.

5.2 Materials and methods

5.2.1 Experimental animals and management

Environmental differences between horses were minimised by selecting subjects from one establishment (Moulton College Equine Yard). This and the availability of horses at the yard resulted in a limitation on the sample size. Subsequently, a sample size of 14 riding school horses (three mares, eleven geldings) was utilised for this study. The age of the horses ranged from five to 18 years (mean age = 12.21 years) and the horses were of various breed types. All horses were regularly exercised (five hours per week) and were managed using the same husbandry techniques and allowed daily access to pasture (eight to twelve hours per day). The aim of the study was to predict behaviour of individuals using subjective ratings of personality, the variation in horse age, breed and gender was not, therefore, considered detrimental to the study.

5.2.2 Horse personality

A regular handler, who had known the horses for a minimum of six months, rated all of the horses using the HPQ but with the five unreliable traits removed (Appendix 8). The HPQ included a worked example and instructions on how to complete the questionnaire. The HPQ was completed for all horses prior to the experimental tests described in Section 5.2.3.

The resulting data were then transformed into scores for the six horse personality components (*Antagonism, Anxiousness, Activity, Protection, Sociability* and *Inquisitiveness*) as described in Section 4.2.3.

Horses were then categorised (high, medium and low) according to their personality scores on each component. The category ranges were determined using the data from the 1223 horses used in Chapter 4 as this provided a larger representation of the range of personalities displayed by horses in general. The categories for a particular personality component were determined by putting all 1223 horses into ascending order according to their scores and dividing the data into three equal groups (see Table 5.2). Thus the range of data within each group was then used to provide the limiters for the horses in this study.

Table 5.2: Category ranges used to categorise personality data.

Component	**Category Ranges** [a]					
	Low		**Medium**		**High**	
Antagonistic	-3.71	-1.38	-1.37	-0.10	-0.09	4.94
Anxiousness	0.27	2.62	2.63	4.05	4.06	9.47
Activity	0.21	5.03	5.04	6.51	6.52	11.10
Protection	0.17	4.85	4.86	6.22	6.23	9.319
Sociability	0.85	5.94	5.95	7.47	4.48	10.63
Inquisitiveness	1.87	7.50	7.51	9.05	9.06	12.62

[a] Ranges calculated using 1223 horses assessed during Chapter 4 and put into ascending order. Low range, 1st to the 408th individual: Medium range 409th to the 818th individual; High range, 819th to the 1223rd individual.

5.2.3 Experimental tests

The experimental tests took place between 3rd April and the 5th May 2006. A one hour buffer period either side of feed times was observed so that no horses were tested when a meal was normally due or had just been consumed. Three tests were used, an arena test, a learning test and a release test, and each horse was assessed once on each test. The tests were designed to measure specific elements of horse personality and were quick and easy to complete.

The arena and learning tests were adapted from Le Scolan *et al.* (1997) and Wolff *et al.* (1997) and had previously demonstrated individual differences in horses. Furthermore, Le

Scolan *et al.* (1997) identified significant associations between the behaviour of horses during these tests and subjective behaviour scores provided earlier, thus demonstrating the validity of these tests and their suitability for this particular study. Furthermore, McCall, *et al.* (2006) demonstrated that isolation of horses from conspecifics was a valid method of assessing reactivity and individual differences.

Arena test

The arena tests were carried out in an indoor arena with sand based flooring. This was a familiar environment to the horses, as they all worked in the arena several times a week. The arena also contained a mirror (4m x 2m) placed on the wall 1m off the floor and at the opposite end to the entrance gate. The mirror was normally present in the arena, thus it was not novel to the horses.

An experienced assistant led each horse to the arena using a standard head-collar and lead rope, except on one occasion where it was necessary to use a bridle and bit. The assistant released the experimental horse upon entry to the arena and an observer recorded its behaviour. The observer was situated centrally down one side of the arena so as to have a clear view of the entire area. The observer remained quiet and still throughout the observation period and was located in a viewing area directly adjacent to the arena. Upon release of the experimental horse, the assistant withdrew to the viewing area of the arena and remained still and quiet throughout the observation period. Both the duration and frequency of the horses' behaviour were recorded using continuous recording methods and all occurrences (Martin & Bateson, 1993). Behaviours were recorded using a check sheet developed from the ethogram shown in Table 5.3. The horses' proximity to the entrance gate was also estimated and recorded (Table 5.3).

Post release into the arena, each horse was observed for ten minutes. This length of time was selected because only the horses' initial reactions to isolation were of interest. In addition, a longer duration may have resulted in horses becoming habituated to the situation or conversely, for those horses that react more negatively towards isolation, may have begun to display higher levels of stress. This would have resulted in poor welfare for the horses. Furthermore, this duration was comparable to the short isolation period (15 minutes) that was used by McCall *et al*. (2006).

Table 5.3: Ethogram of horse behaviours recorded during the arena test

Behaviour	Definition
Gate	Horse positioned within 5m of the entrance gate
Active locomotion	Movement in any gait including: walk = move forward with slow four beat gait (Strand *et al.,* 2002), horse walks energetically, looks in front of self and around (Le Scolan *et al*., 2002); trot = two beat gait (Strand *et al*., 2002); canter = three beat gait (Strand *et al*., 2002); gallop = fast four beat gait (Strand *et al*., 2002); passage = animated form of trot where the legs are raised with more elevation often associated with audible hoof contact with the ground (Le Scolan, *et al*., 1997)
Exploration	Walks slowly with neck horizontally or lower, ready to stop and sniff the ground or other object, characteristic walk of a quiet horse in a calm situation (Le Scolan *et al*., 1997)
Vigilance	Stand with eyes focused forward for 5s or more (Strand *et al.*, 2002). Elevated neck intently orientated head and ears (Le Scolan *et al*., 1997) Ears held stiffly upright, nostrils dilated (McDonnell & Haviland, 1995)
Vocalisation	Any vocalisations made by the horse, whinnies, neighs etc.
Spook	Move abruptly in any direction in a manner typical of avoidance or removal from an area (Strand *et al.*, 2002)
Stand	Stand with eyes either down or forward, ears soft for 5s or more. (Strand *et al.*, 2002)

Learning test

The learning task required the horse to open an adapted feed trough with a lid in order to access food placed within the trough (Figures 5.1 and 5.2). The trough measured 42 x 39 x 23.5 cm and was closed by a multi density fibre (MDF) lid with two hinges and an over-hanging lip of two centimetres at the side opposite the hinges. The lid also had several small holes (0.5 cm diameter) in order to allow the horse to smell the food inside. In order to weigh the trough down, it was securely wedged into a large tyre (Dunlop, 195/61,R15 88H), to prevent the horse from accessing the food by simply knocking the trough over. The trough and tyre were placed on the floor of the horse's stable, in the corner directly opposite the doorway, where it could be clearly viewed by the observer.

The observer recorded the following behaviours; latency to first touch the trough (LT) and time taken to successfully open the trough (LTO) and access the food. Additionally during each phase at ten second intervals it was recorded whether the horse was touching the box (Total T) this was used as a measure of interest in the trough. The predicted links between these measures with the specific personality components are identified in Table 5.1. The test involved the following three phases and was an adaptation of the studies carried out by Le Scolan *et al.* (1997) and Wolff *et al.* (1997). Each phase lasted three minutes so as to ensure that horses were interested and motivated in attempting to open the trough.

Phase one: An assistant restrained the experimental horse in its home stable using a standard head-collar and lead rope. The observer then placed the trough in the stable and demonstrated the opening mechanism to the horse and placed half a litre of food (Dodson and Horrel Pasture Nuts) into the trough in full view of the horse and then closed the lid. Once the demonstration was complete, the horse was released for the three minute duration of the test. Both the observer and the assistant left the stable and stood two metres outside

the stable door and observed quietly. The observer used a recording sheet to record LT, LTO and Total T as defined earlier.

Phase two: If the experimental horse had not successfully opened the box at the end of the first three-minute trial then the assistant would restrain the horse as before and an additional demonstration was provided. In addition the observer held out a handful of food toward the horse such that it could smell but not eat it. The horse was then released to complete its second three-minute trial and its behaviour recorded as in phase one.

Phase three: If at the end of the second trial the horse still had not been successful in accessing the food, the horse was again restrained and the demonstration repeated by the observer. This time the horse was provided with a small mouthful of the food and was then released to complete its final trial and its behaviour recorded as in phases one and two.

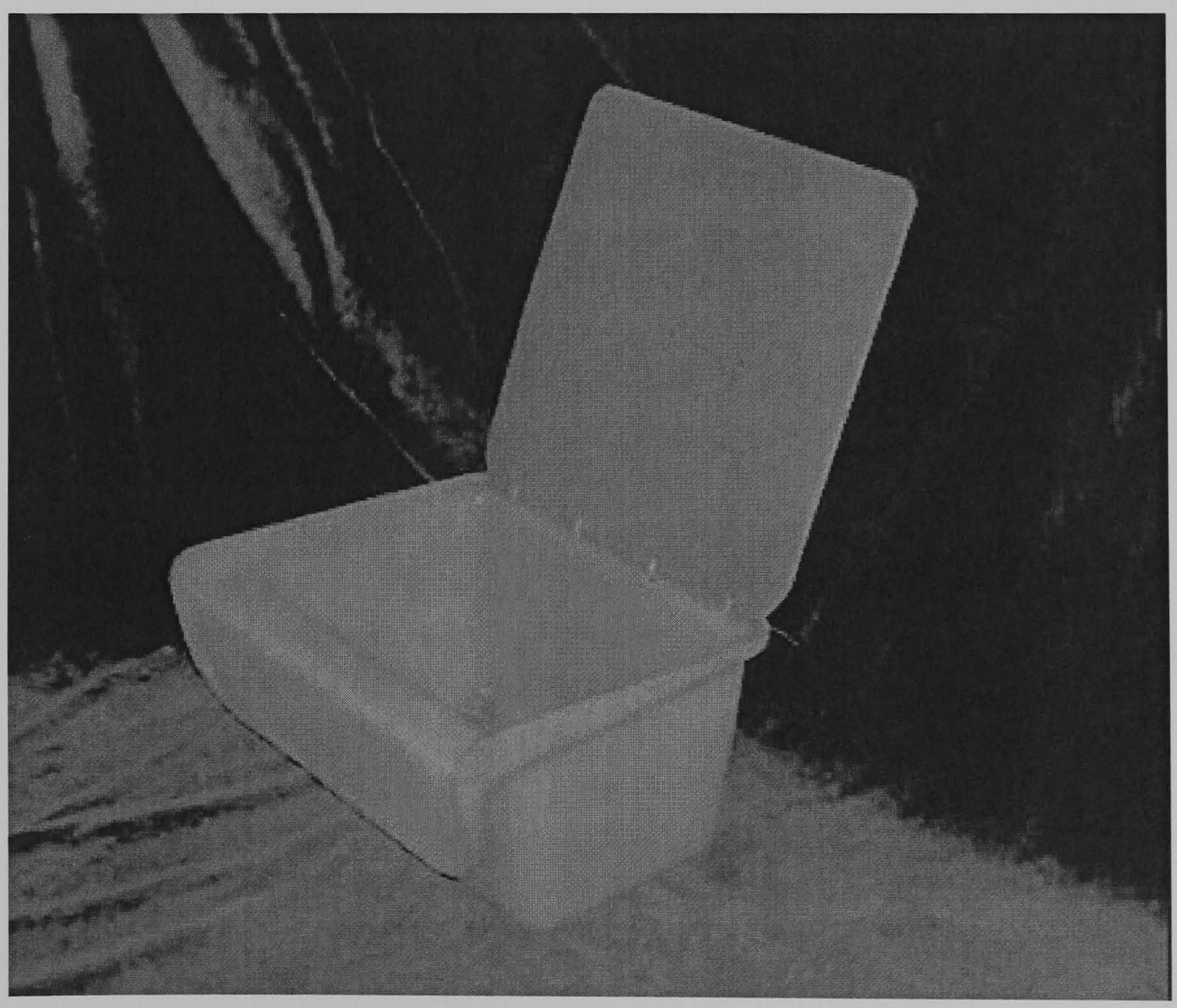


Figure 5.1: Adapted feed trough with lid (open), used during the learning test

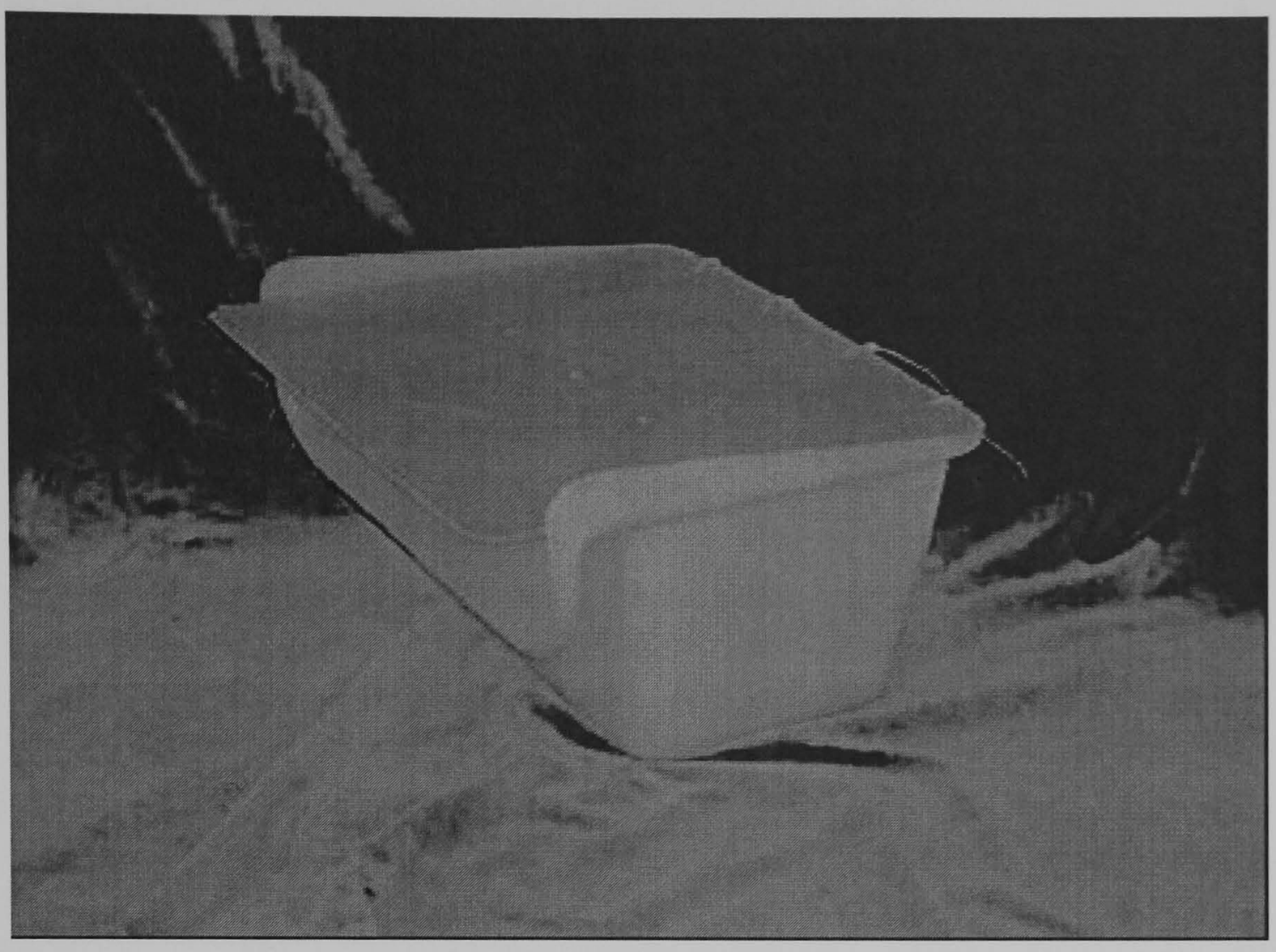


Figure 5.2: Adapted feed trough with lid (closed), used during the learning test

Release test

The assessments took place during the horses' usual turnout routine. Horses were led to their usual paddock, using a standard head-collar (with the exception of one horse, where a bridle was used) and a familiar handler. They were observed whilst being led to the paddock and during the first ten minutes after release. At least two of the focal horse's usual conspecifics were already in the paddock prior to release, and had been for at least ten minutes.

On the approach to the paddock the horse was scored on a scale of one to four on activity and ease of handling (Table 5.4). The scale was developed through discussions with horse handlers with regards to the general behaviours displayed by horses whilst being led to the field and the resulting reactions displayed by the handlers. Preliminary observations of horses being led to their field also aided the development of the scale. The measurement scale was designed to provide a quick and simple way of rating a horse on its handleability. This assessment method is similar to that used by McCall *et al.* (2006) in their assessment of horse emotionality and reactivity.

Table 5.4: Definitions of ease of handling ratings given to horses on approach to paddock during the release test.

Ease of handling rating	Horse performs at least one of the following behaviours
1 (very easy)	Slow walk – head down/ low Handler relaxed and holds rope with relaxed grip Horse appears relaxed
2 (easy)	Horse appears alert Walk appears hurried with elevated head Horse looking around – head carried high Handler generally relaxed but with firm grip
3 (difficult)	Horse trots (two beat gait) – may be combined with walking Handler tense with firm and short grip of rope – probably needing to restrain horse Horse very alert – head carried high May show low levels of aggression – e.g. ears back, napping, threats only
4 (very difficult)	Performs any of the following; Bolt, Spin, Rear, Bite, Kick

On arrival at the paddock, the horse was released to join its conspecifics and its behaviour recorded during the following ten minutes. As with the arena test, a ten minute observation period was selected in order to provide an indication of the horses' initial reactions to being released into the field. The time taken for the focal horse to approach another individual (movement directed towards a specific individual and the focal horse stops at least five metres from another individual) was recorded and the amount of antagonistic and affiliative behaviours both received and given were recorded (Table 5.4). Antagonistic encounters included either performing or receiving a kick, kick threat, bite, bite threat, head threat, herding or chasing behaviours. Whereas affiliative interactions included performing or receiving behaviours such as, mutual grooming, greeting, playing and grazing within close proximity (less than two metres) to another individual (see Table 3.2 for definitions of each behaviour).

Statistical analysis

Data were analysed using SPSS Version 14 for Windows (SPSS Inc., USA). Prior to analysis several predictions were made as to the expected differences in behaviour across personality categories (i.e. high, medium, low) for specific personality factors (Table 5.1). The predictions were tested using the Kruskal-Wallis test and tested for differences in behaviour between the personality categories described in Section 5.2.2. Alpha was set at 0.05 for all statistical tests.

5.2.4 Ethics and welfare

The methodology for this experiment was assessed and approved by the Moulton College Research Committee. The assessment of the horses using the HPQ was carried out by a regular handler of the horses and did not require any manipulation of the horses and therefore had no negative impact on horse welfare.

In contrast the behaviour tests described in Section 5.2.3 did involve some direct manipulation of the horses. The release test was thought to have had the least impact, in that there was no significant deviation from the horses' normal routines. They were turned out by familiar handlers with their usual cohort and were released at the usual time of day.

Similarly the arena test had a minimal impact on the horses as they were being released into a familiar area. Although they were not being handled by a familiar person, this was not unusual for these horses as they are based at an equine college. It was therefore common for these horses to be handled by unfamiliar people. The handler used in this particular experiment had at least 20 years experience of handling horses and was therefore highly competent thus minimising any stress during handling.

The purpose of the arena test was to isolate the horses from their conspecifics for a short period of time in order to assess their reaction to social isolation. Although this may have resulted in the horses becoming stressed it was only for a short period of time and was thought to be no worse than the common practice of exercising horses away from conspecifics.

The main welfare concern in the learning test was ensuring that the equipment used would not pose any risk to the horses being assessed. The modified feed trough used in the experiments was originally designed for use with horses and any additional modifications were done so in such a way as not to have any sharp edges that would otherwise have posed an injury risk. The use of a tyre to weigh down the feed trough meant that the horses would not be able to significantly move it (e.g. pick up and throw it), thus further reducing the risk of injury. In addition the horses were also monitored for the full period that the equipment was left in the stable. This meant that if a horse became aggressive or frustrated with the box, then the observers could restrain the horse and remove the equipment, thus further reducing the risk of injury.

The other major consideration of the learning test was the use of food as the reward. The food used was part of the horses' normal diet and was selected through consultation with the horses' carers as was the volume of food used. As such the use of the food reward did not pose any welfare issues.

Overall it was determined by the author and the Moulton College Research Committee that these experiments were ethical and did not negatively affect the welfare of the horses involved in the study.

5.3 Results

5.3.1 Horse personality

Horses demonstrated individual differences and the personality component scores for all 14 horses are shown in Table 5.5.

5.3.2 Arena test

Horses reacted in range of different ways during the arena test. See Table 5.5 for a summary of behaviours recorded. No significant differences were identified between any of the personality components and their predicted behaviours (Table 5.6)

5.3.3 Learning test

A total of seven horses (50%) successfully opened the trough and accessed the food. Of those, three managed to open the trough in the first trial, three needed the second trial and one required all three trials. Only one horse did not touch the trough, the remaining 13 horses made first contact within 40 seconds (n = 13, mean = 9.62 seconds, SD = 12.19) (see Table 5.5). Data from all 14 horses were entered into the analysis. No significant differences were identified between any of the personality components and their predicted behaviours (Table 5.6)

5.3.4 Release test

Of the 14 horses, 12 were rated as being easy to handle (handling score ≤ 2) the highest handling score given was three (n = 2) and none were rated as level four. See Table 5.5 for a summary of results. No significant differences were identified between personality categories and the predicted behaviours (see Table 5.6).

Table 5.5: Personality scores and behaviour data for all 14 horses involved in the prediction study

Variable	Horses														Mean	SD
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**12.21**	**5.25**
Age	16	5	18	5	5	19	16	17	12	17	8	12	14	7		
Gender	F	M	M	F	M	M	F	M	M	M	M	M	M	M	0.38	1.36
Antagonism	-0.11	-0.16	0.08	0.94	-0.92	2.68	2.66	-0.34	-0.62	-1.30	0.45	-1.02	2.53	0.42	4.56	1.66
Anxiousness	4.37	1.69	3.03	3.79	5.45	3.45	3.62	2.83	6.39	5.46	7.11	6.46	3.69	6.45	5.97	1.65
Activity	3.53	2.78	4.84	4.99	4.69	7.78	7.05	5.65	7.66	5.95	6.51	6.28	7.80	8.06	4.56	1.91
Sociability	2.85	1.20	5.52	3.93	4.68	3.78	1.44	5.04	7.85	5.60	4.95	6.16	3.75	7.04	4.51	2.21
Protection	3.21	1.80	4.72	3.15	3.95	2.26	1.16	6.23	7.57	6.86	4.34	6.76	3.29	7.88	8.24	1.72
Inquisitiveness	6.00	6.14	8.98	9.47	6.76	10.66	6.28	9.28	9.38	7.15	7.74	6.90	9.72	10.90	47.50	142.23
LT [a]	4	5	4	2	6	38	2	5	35	1	10	8	5	540	393.21	193.17
LTO [a]	540	346	540	540	325	300	523	540	540	540	44	82	105	540	18.43	15.03
Total T [b]	20	15	2	49	18	35	32	8	40	14	5	9	11	0	16.21	22.92
Gate [a]	54	17	0	0	0	13	25	0	0	66	46	0	0	6	86.21	62.56
Active locomotion [a]	178	74	66	61	22	53	101	106	14	233	16	122	115	46	356.36	106.73
Exploration [a]	362	376	296	253	516	368	349	236	467	129	462	398	304	473	46.57	59.83
Vigilance [a]	0	88	0	188	25	161	22	6	24	0	61	34	35	8	99.71	79.30
Stand total [a]	60	62	221	98	19	9	128	239	82	228	50	25	132	43	0.93	1.21
Spook [b]	0	0	2	0	0	0	0	0	2	0	3	1	3	2	2.71	8.30
Vocalisation [b]	6	0	0	0	0	1	0	0	0	31	0	0	0	0	44.64	43.16
LTA [a]	10	64	16	79	10	5	10	69	86	60	155	28	10	23	2.50	2.07
Affiliative [b]	4	2	3	2	4	1	3	5	0	0	3	1	0	7	3.79	4.81
Antagonistic [b]	3	1	5	5	19	3	6	0	0	1	2	5	2	1	1.50	0.76
EoH [c]	1	1	1	2	2	1	1	1	1	1	3	3	2	1	12.21	5.25

LT, latency to touch; LTO, latency to open; Total T, time touching box; EoH, ease of handling; LTA, latency to approach; Affiliative, number of affiliative interactions given and received; Antagonistic, number of antagonistic interactions given and received; [a] duration (measured in seconds); [b] , frequency data; [c] rating data; M, castrated male (gelding) F, female (mare)

Table 5.6: Kruskal Wallis test on behaviour using personality categories (i.e. high, medium and low) as the grouping variables.

Test	Behaviour	Kruskal Wallis (χ^2)			
		Anxiousness	***Activity***	***Sociability***	***Inquisitiveness***
Arena	Gate	1.68	*n/a*	1.92	*n/a*
	Active locomotion	0.32	0.00	0.36	*n/a*
	Exploration	*n/a*	*n/a*	*n/a*	2.08
	Vigilance	5.27	*n/a*	3.61	*n/a*
	Vocalisation	0.74	*n/a*	0.60	*n/a*
	Spook	2.23	1.60	*n/a*	*n/a*
	Stand	2.39	1.44	2.39	*n/a*
Learning	LT	*n/a*	*n/a*	*n/a*	0.89
	LTO	*n/a*	2.25	*n/a*	0.78
	Total T	*n/a*	*n/a*	*n/a*	3.64
Release	EoH	*n/a*	0.11	*n/a*	*n/a*
	LTA	0.49	*n/a*	2.13	*n/a*
	Affiliative	*n/a*	*n/a*	4.62	*n/a*
	Antagonistic	*n/a*	*n/a*	5.29	*n/a*

All χ^2 values non-significant; *n/a* = not applicable; LT, latency to touch; LTO, latency to open; Total T, time touching box; EoH, ease of handling; LTA, latency to approach; Affiliative, number of affiliative interactions; Antagonistic, number of antagonistic interactions.

5.4 Discussion

The results of this study failed to identify any significant differences in behaviours between personality categories thus suggesting that the results of the HPQ were not able to predict a horse's behaviour during the selected behaviour tests. These results indicated, therefore, that the HPQ does not meet Gosling and Vazire's (2002) Criterion Two (assessments are predictive of behaviours and real-world outcomes). The overall predictive use of personality scores does, however, require further investigation. There are two main hypotheses for the inability for the HPQ to predict behaviour in this study: 1) the HPQ is not appropriate for the prediction of horse behaviour, or is not assessing personality with sufficient accuracy; and 2) the behavioural tests and/or the behaviours measured were unsuitable for comparison to personality scores due to their specificity and time constraints. The following discussion explores the possible reasons for the HPQ being unable to predict behaviour.

It was predicted that *Sociability* would be associated with the behaviours observed during the arena and the release tests. As horses are gregarious (McDonnell, 2002) isolation from conspecifics should be relatively stressful experience for horses and initiate a variety of behaviours such as vocalisation and increased activity. The level of stress exhibited would be expected to vary with personality and in particular with an individual's score on *Sociability*. Although there were a variety of reactions performed by the horses during both the arena and the release tests none of these behaviours were found to be predicted by *Sociability*. It is possible that the horses used in this study had become acclimatised to working in the arena in the absence of other horses. They may not, therefore, have differentiated the arena test as being anything different from routine and perhaps accepted that they would soon be returned to their stable and their conspecifics. Furthermore,

Sociability is defined by the trait playful, which was previously found to be associated with play behaviour in horses (Tables 3.4 and 3.5) and affinitive behaviours (e.g. approach) in primates (Capitanio, 1999; Pederson *et al.*, 2005). Play behaviour was not directly assessed in any of the tests and may partly explain the lack of links between *Sociability* and behaviour.

Excitability (defined by the traits: active, slow (-), excitable and intelligent), was predicted to be linked with an individual's level of activity in the arena test, ability to learn a simple instrumental task and ease of handling in the release test. *Excitability* was not however, significantly predictive of any of the behaviours. Levels of activity varied greatly between the horses but were not significantly associated with *Excitability*. In horses at grass, time spent standing has been shown to be negatively correlated with *Excitability* (Tables 3.4 and 3.5). It was therefore surprising that in this experiment time spent standing could not be predicted by *Excitability* scores. It is possible that as a result of the short observation time in the arena test and the restricted number of horses available, that insufficient data were collected in order to demonstrate this link. A more extensive study would allow for more detailed analysis that may allow for individual locomotor behaviours such as canter and trot to be analysed with respect to *Excitability*. As personality is expressed in detailed and complicated ways, more detailed measurement of behaviours may be required.

The personality component *Anxiousness* was previously shown to be associated with the occurrence of passage (Tables 3.4 and 3.5), but was not found to be predictive of any of the activity measures in this study. It is possible that through the grouping of similar behaviours, important detail was lost with regards to the animals' behaviour, which in turn made comparison to personality scores more difficult. *Anxiousness* was, expected to be predictive of spook frequency in the arena test but this was not demonstrated. Momozawa

et al. (2003) previously identified that *Anxiousness* was significantly associated with fear responses, for example, increase in heart rate and defecation frequency in horses exposed to balloon reactivity tests. Such novel object tests may, therefore link more strongly to ratings of *Anxiousness*.

Inquisitiveness was expected to predict the results from the learning test and levels of exploration in the arena test. Horses showed a range of skills and approaches to opening the trough, but only half managed to successfully access the food. Of the seven unsuccessful horses, most lost interest in the trough after the first few minutes of each phase of the test, demonstrating varying levels of determination. Despite this apparent range of individual differences, *Inquisitiveness* was unable to predict a horse's performance in this test. *Inquisitiveness* is a measure of how opportunistic and curious a horse is thought to be, and may not be as predictive of learning ability as first thought. A novel object test may therefore be more appropriate for assessing *Inquisitiveness*, although the influence of fear would need to be assessed and quantified.

The reliability and validity of the HPQ was assessed in earlier studies and demonstrated good inter-rater agreement and links to observed behaviour (Chapter 3) as well as personality differences between breeds and providing support for anecdotal evidence of breed typical behaviours (Chapter 4). Other rating methods similar to the HPQ have proven to be reliable for the assessment of other animal species (for example, Stevenson-Hinde & Zunz, 1978; Stevenson-Hinde *et al.*, 1980; Figueredo *et al.*, 1995; Capitanio *et al.*, 1999; Creighton, 2003; Martin, 2005) as well as demonstrating links with observed behaviour (Capitanio, 1999; Wielebnowski, 1999; Momozawa *et al.*, 2003; Pederson *et al.*, 2005) and biological measures (Capitanio *et al.*, 1999; Momozawa *et al.*, 2003; Capitanio *et al.*, 2004). These studies provided evidence that supports the use of rating questionnaires

as a reliable method of personality assessment in animals and indicated that prediction of behaviour from personality is possible (Capitanio, 1999; Pederson *et al.*, 2005). The use of the HPQ for the assessment of horse personality in the present study was therefore valid and appropriate, and does not explain the absence of links between personality and behaviour in this study. Focus should therefore be directed towards assessing the validity of the behavioural tests and measures recorded.

The behavioural tests employed in this study demonstrated individual differences in horse behaviour, but few links between personality scores and behaviour were identified. Both the arena test and the learning test used in this study were adapted from Le Scolan *et al.* (1997) who assessed the behaviour of 72 horses from three riding schools. Their study also compared behavioural results with a simple rating questionnaire, which rated the animals on a scale from one to three on seven items, for example, 'fearful when ridden' and 'socially dependent in an unknown surrounding'. The results identified correlations between some of the rating items and measured behaviours. For example, horses rated as 'good learners' needed less time to open the chest containing the food than those rated as 'bad learners', indicating that the learning test was a good predictor of more general learning abilities. Also the more 'socially dependent' a horse was considered to be the more reactive the animal was in the arena test. Individuals with high emotionality indices (calculated from behaviour recordings in the arena test) were also rated as highly gregarious and such horses were difficult to separate from conspecifics in working situations. The arena and the learning tests have therefore previously demonstrated individual differences and personality-behaviour correlations in horses. These tests are therefore valid and appropriate for directly measuring behavioural differences in horses and these behaviours should be predictable from the results of the HPQ.

The use of questionnaires and behavioural tests have previously been shown to be valid for the assessment of individual differences in horses and yet their results were not comparable. Visser *et al.* (2003a) had similar difficulties when trying to predict show-jumping performance in horses. The behavioural tests they employed were unable to accurately predict performance. Visser *et al.* (2003a) hypothesised that their tests were not assessing personality traits that were relevant to show-jumping performance and that further development of the tests would be required in order to make more accurate predictions. It seems likely that this may also be the case in the present study. The behaviour tests used were assumed to be assessing certain aspects of horse personality and that the behaviours recorded would reflect these personality characteristics. It is possible that the behaviours recorded were not relevant or were restricted by sample size and were therefore insufficient to identify personality-behaviour correlations. It is therefore likely that it is not the HPQ that is failing to predict behaviour, but it is the human interpretation of how personality and behaviour should correlate that may need reassessing and future studies should address this.

5.5 Summary

The results of this study failed to demonstrate that personality scores calculated using the HPQ could predict the behaviour of horses in specific situations. These results indicate that these assessment and testing methods require further development in order to make more accurate predictions. Further research is required to explore the personality-behaviour links that exist in horses in order to further understand which personality characteristics are most likely to be predictive of future behaviour. This is required to meet Gosling and Vazire's (2002) Criterion Two that personality assessment methods must be able to "*predict behaviour and real-world outcomes*" and further validate the HPQ as a method of personality assessment. This is of great importance if the HPQ is to be incorporated into selection processes in equine disciplines of studbook registration, where assessments must be valid and reliable.

6 Discussion

6.1 Project summary

This project has demonstrated that a novel method of horse personality assessment (the HPQ) is reliable, easy to use and can be linked to real-world outcomes. Furthermore the results of the three experiments generate questions that can be incorporated into future research in order to further develop our understanding of animal, and more specifically, horse personality.

Firstly, in Chapter 3, the Stevenson-Hinde and *et al*. (1980) trait list was adapted, thus creating the HPQ. The use of this rating method was then shown to be reliable with a high inter-rater reliability. Furthermore, significant correlations were shown between horse personality scores and behaviour recorded whilst at grass. It was therefore concluded that the horse personality data met both Criteria One (independent assessments must agree) and Three (ratings must reflect genuine attributes of the individual) and were indicative of Criterion Two (assessments must be predictive of behaviour and real-world outcomes) (Gosling & Vazire, 2002). In addition a six-component structure of horse personality was identified using PCA. The resulting components were labelled *Antagonism, Activity, Anxiousness, Protection*, *Sociability* and *Inquisitiveness* and explained 76.53 % of the total variance in the data. These results were shown to be comparable to those found in similar studies on both equines and non-equines and provide strong evidence for the existence of horse personality and our ability to assess it reliably.

In Chapter 4 the HPQ was further validated through the exploration of breed typical personality. Over 1200 horses, from eight different breeds were assessed using the HPQ and the data analysed for breed differences in personality. The results identified strong inter-breed differences, with *Anxiousness* and *Activity* having the greatest inter-breed variation. Furthermore, those breeds with linked pedigrees or functions were shown to have similar personality types. The results of this study provided further evidence that the HPQ meets Gosling and Vazire's (2002) Criterion Two, through the demonstration of links with real-world outcomes, which in this case were breed differences that can be explained in terms of breed function.

Finally, Chapter 5 focused on testing whether an individual's personality rating could be used to predict future behaviour in a series of behaviour tests. This was to further ensure that the assessment method met Gosling and Vazire's (2002) Criterion Two and therefore demonstrate the validity of the HPQ. Fourteen horses were each assessed using the HPQ and then exposed to three behavioural tests. None of the predicted links between behaviour and personality were identified. Having previously demonstrated the reliability of the HPQ, it was proposed that the behavioural tests or measurements employed were not suitable for being linked to personality assessments. Furthermore, personality-behaviour relationships may not be as simple as were originally predicted, and require further investigation.

In summary, all three criteria set out by Gosling and Vazire (2002) have been met by this project and the results, therefore, demonstrate that the HPQ is a reliable and valid assessment method for horse personality. The suitability of the HPQ for the prediction of behaviour, however, requires further investigation. The project will now be discussed in

terms of methodology, implications for personality research and the potential applications of the HPQ within the equine industry.

6.2 Methodological approach

In Section 2.1 the different approaches to personality assessment were discussed and trait-rating and behavioural assessments were highlighted as the two main approaches used in animal personality assessment. In order to demonstrate validity, these assessments must be shown to meet three criteria (Gosling & Vazire, 2002) (see Section 2.2). The HPQ has been shown to meet these criteria and, therefore, indicated that the use of questionnaire assessment for personality is both reliable and valid for horse personality. In addition, the reliability of the data inferred that the HPQ meets the 'good trait' criterion described by Funder (1995). It has therefore been demonstrated that, with only minor alterations, the Stevenson-Hinde *et al.* (1980) trait list can be successfully applied to horses.

In Chapter 3, rater reliability was shown to be high, with only 17 (27.9%) horses not rated reliably. This further emphasises the importance of testing for rater reliability in animal personality studies. Using mean scores between raters (e.g. Momozawa *et al.*, 2003) without first ensuring that raters are in close agreement, is likely to produce inaccurate personality ratings. Although 100% agreement between raters may be unlikely, it is important that it is quantified.

The 17 horses that were not rated reliably in Chapter 3 were described as being 'bad targets', such that they were considered more difficult to rate than the other horses (Funder, 1999). Level of acquaintance between raters and targets has been shown to affect human personality ratings (Funder *et al.*, 1995), it is likely, therefore that level and type of acquaintance will affect the reliability of animal personality assessments. For example, regular handlers may work on a daily basis with those animals being assessed; nevertheless

their exposure to an animal's behaviour will vary depending on the contexts in which they see the individual.

Moreover, Funder (1995) explores the concept of 'good information' in that more or certain kinds of information may make for more accurate judgements. The availability of such information should therefore be considered when deciding on the suitability of particular individuals as raters. It is also possible that some individual horses may react differently to different handlers. This may vary depending on the animal's level of trust or experience with each handler, and is a further expression of its personality, but one that may be difficult to assess. Such variables require further investigation in order to improve reliability and to determine which type of handlers are the most appropriate for rating horse personality, i.e. it is important to identify what makes a 'good judge' (Funder, 1995). Due to the number of different equine establishments used in this study and the lack of consistent raters across all horses, it was not feasible to explore the variables that may affect personality ratings by horse handlers. Future studies on horse personality would need to specifically explore these variables. Similar focus on the variables affecting rater reliability was also suggested by Gosling and Vazire (2002) with respect to general animal personality research.

To further demonstrate the reliability and validity of the HPQ, the assessments needed to be predictive of behaviour and real-world outcomes (Criterion Two) as well as demonstrate that raters were providing genuine assessments of personality (Criterion Three). In Chapter 3, personality scores were shown to significantly correlate with some of the recorded behaviours. These results indicated that the ratings provided by the regular handlers were genuine assessments of the horses' personalities, demonstrating that Criterion Three had been successfully met.

The links with observed behaviour identified in Chapter 3 inferred that personality assessments could be used to predict horse behaviour. When this was tested in Chapter 5, the results were disappointing as none of the predicted personality-behaviour difference were significant. As the HPQ has been demonstrated as a reliable method of assessment and links with behaviour have been identified, it seemed most likely that the behaviour tests were not appropriate for the prediction of behaviour from personality scores. The limitations of behavioural tests were discussed in Section 2.1.4. Although capable of identifying individual differences in behaviour, they are often restricted in the elements of personality that they can reliably assess. The difficulty in measuring social aspects of a horse's typical behaviour means that current behaviour tests are in fact restricted to measurement of temperament as opposed to personality as a whole. It is also possible that the short duration of the tests only provides a 'snap-shot' of the horse's personality and may be insufficient to show significant correlations with personality ratings. Perhaps a more appropriate test for the predictive use of the HPQ would be to predict the success of horses entering a training programme. For example, young horses selected for police training could be assessed with the HPQ at the beginning of the training process. Horses with specific personality types would be predicted to perform better throughout the training and successfully complete the programme. For example, it might be predicted that successful horses would score low on *Anxiousness* and *Excitability* but may score highly on *Inquisitiveness*. Testing the HPQ in this way would also begin the process of adapting it for use in the equine industry.

6.3 Implications for personality research

Having demonstrated that the HPQ meets all three of Gosling and Vazire's (2002) criteria, these results provide further evidence for the existence of animal personality and our ability to measure it accurately. Furthermore, it has been demonstrated that an assessment tool developed for rhesus macaques can be successfully adapted for use with horses. Consequently, these results can be used in cross species comparisons.

6.3.1 Breed differences and the heredity of personality

The significant breed differences in personality identified in Chapter 4 provide strong indication that humans have artificially selected horses on both morphology and behaviour (i.e. personality). Few studies have investigated breed differences in horse behaviour. Hausberger *et al.* (2004) indicated that horses of different breeds react differently to a bridge test, therefore indicating that some inter-breed differences in behaviour and reactivity exist. Moreover, anecdotal evidence from breed societies indicated towards the existence of breed typical behaviours. The results of the second experiment (Chapter 4) clearly identified breed differences in personality. Additionally, similarities between breeds with linked pedigrees and functions were identified. Breed differences in personality were most common on the components *Anxiousness* and *Activity*, resulting in a gradient of personality types on these components. For example, those horses bred for speed, ranked highly on these components in comparison to those bred for draught work and strength. As the breed differences were so strong between unrelated breeds and the similarities between linked breeds consistent across components, this study provides further evidence that personality is heritable and specific traits can be selected for. To horse breeders and owners this is not a novel concept, but one that science is slowly beginning to catch up with. The evidence for heritability of personality is increasing (e.g.

Weiss *et al.*, 2000; Bouchard & Loehlin, 2001; Gauly *et al.*, 2001; Dingemanse *et al.*, 2002; Drent *et al.*, 2002; van Oers *et al.*, 2003; Fairbanks *et al.*, 2004) and is now being linked to theories of personality evolution (Buss, 1991; Bouchard & Loehlin, 2001; Dall *et al.*, 2004; Sih *et al.*, 2004a).

Nevertheless, it should be noted that personality, although heritable to some degree, is a result of a combination of environmental and genetic influences (Archer *et al.*, 2003; Stamps, 2003; Sih *et al.*, 2004b). Although not explored in this study, the environmental effects on horse personality would be of great interest. For example, can specific environmental conditions be detrimental to a horse's personality? It is already known that the welfare and behaviour of horses can be affected by environmental conditions (e.g. Heird *et al.*, 1986; Redbo *et al.*, 1998; Cooper *et al.*, 2000; Christensen *et al.*, 2002; Heleski *et al.*, 2002; McAfee *et al.*, 2002; Rivera *et al.*, 2002; Chaya *et al.*, 2006; Ninomiya *et al.*, in press), but the effects of early environment on the development of horse personality has not been followed. This would be a challenging longitudinal study that may have significant welfare implications and is worthy of further research.

6.4 Why do horses have personalities?

Personality develops through a combination of genetics and environmental input (Zuckerman, 1991a; Archer *et al.*, 2003). Archer *et al.* (2003) demonstrated that even when genetic variation is controlled for, individual differences can still exist. Although genetics clearly have a strong influence on personality (e.g. Zuckerman, 1991a; Weiss *et al.*, 2000; van Oers *et al.*, 2004; van Oers *et al.*, 2005), the environment an individual develops in will also have a significant influence (Zuckerman, 1991a; Stamps, 2003). The combination of these variables results in a broad spectrum of possible outcomes with regards to an individual's behaviour and personality, just as it does for morphological features. It is no wonder then that individual differences exist in a wide variety, if not all, animal taxa.

Until recently intraspecific differences in behaviour were seen as non-adaptive variation surrounding (possibly) adaptive advantages (Dall *et al.*, 2004). The non-random manner in which inter-individual variation in behaviour is often distributed along particular axes (Gosling & John, 1999; Gosling, 2001), suggests that personality is likely to have consistent ecological and evolutionary consequences and therefore would be a focus for selection (Dall *et al.*, 2004). Personality differences are likely to be selected in combination with morphological characteristics. This co-evolution may facilitate the conservation of adaptive changes and speciation in response to environmental change (Dall *et al.*, 2004). Recent reviews of personality genetics and evolution are beginning to demonstrate that consistent individual differences in behaviour can have adaptive advantages and have been selected for by natural selection (Bouchard & Loehlin, 2001; Dall *et al.*, 2004; Sih *et al.*, 2004a; van Oers *et al.*, 2005). In domesticated species these selection processes have also involved anthropogenic (artificial) selection.

Horses are gregarious animals that would naturally live in social groups consisting of a dominant male with several females and their young (Clutton-Brock, 1999). Young males leave their natal herd to gather in bachelor groups until they are capable of forming their own breeding herds (Clutton-Brock, 1999). In such a social environment it would be beneficial to have an idea of how other members of the group typically behave. If behaviour is consistent across situations then interactions between individuals may become less costly as each individual can make an informed decision on how to act with each member of the group. For example, the potential for costly interactions, such as fighting over resources, would be reduced as those individuals that are more aggressive, would be less likely to be challenged by others that are less aggressive; this would then result in a dominance hierarchy. The potential mechanisms involved in the evolution of personality within populations are reviewed by Dall *et al.* (2004). They discussed the advantages of being able to predict an opponent's behaviour and suggest that personality should be analysed in terms of a dynamic and state-dependent game theory. The existence of personality basically provides other individuals with an opportunity to minimise the cost of interactions, and may be of greatest advantage to gregarious species. It would be interesting to compare personality structures and scores between gregarious and more solitary species, as it may be that personality may be more prominent and multidimensional in social species. This may be best carried out using primate species due to their morphological and behavioural similarities and yet varying social systems.

The domestication of animals is seen as strong evidence for the genetic influence on behaviour (Bouchard & Loehlin, 2001). Throughout the process of domestication, humans have selected for those individuals whose behaviours allow for easy control and management (Clutton-Brock, 1999). Further selection of animals for specific purposes results in the formation of different breeds, which vary not only in morphological features,

but also in their behaviours (Clutton-Brock, 1999). Until recently, few studies have explored inter-breed differences in behaviour. Most breed studies to date have focused on dog breeds and have used both behavioural (Serpell & Hsu, 2005; Svartberg, 2006) and questionnaire assessments (Bradshaw *et al.*, 1996; Notari & Goodwin, in press) to investigate any inter-breed differences. These studies have been able to demonstrate clear differences between breeds (Bradshaw *et al.*, 1996; Notari & Goodwin, in press) and links between personality and breed function (Svartberg, 2006). Extensive comparisons of horse breeds had not previously been explored. The evidence from Chapter 4 suggests that artificial selection on horse breeds has had a substantial effect on the typical behaviours of the breeds studied. This selection is still ongoing, with some breed societies choosing to use temperament tests during the grading of stallions and brood mares (e.g. Franches montagne, Mischel, June 2006, Personal communication). Those horses with undesirable temperament traits that would make them difficult to handle, such as high levels of aggression, are removed from breeding. Although these selection processes are important to the maintenance of breeds, the original function and character of the breeds should be kept in mind. As Svartberg (2006) demonstrated, modern selection for show ring traits can affect the overall personality type of a breed and deviate personality trends from those originally sought. Such characteristics are likely to be different to those originally selected when the animals were bred for work. Breed societies may need to consider whether these characteristics should be conserved with the same emphasis as morphological characteristics such as size and coat colour.

From a species perspective personality is adaptive, and with respect to horses it may have aided their survival as a social species. More recently, these variations have been utilised to human advantage as opposed to species survival.

6.5 Applications and future directions

The potential applications of horse personality assessment were discussed in Chapter 1. The main application of the HPQ would be to incorporate it into selection processes for specific equine disciplines. The use of personality assessment, using behavioural tests, has already been explored by Visser *et al.* (2003a) for the prediction of show jumping performance in horses. Their results were, however, inconclusive and failed to show any strong predictive qualities as has also been demonstrated in this study. This may have been due to the specificity of the behavioural tests. The HPQ is, however a more holistic approach to horse personality assessment and may be better suited to predictions of performance over time rather than specific behaviour events.

The HPQ does, however, require further development and testing prior to application within the industry. Rater reliability would require further research, in order to identify those factors that make individuals good raters. The predictive potential of the HPQ would also require further investigation. This may be best demonstrated by predicting industry relevant outcomes, for example, success in a training programme. As opposed to behaviour tests that are both time consuming and restrictive and generally not directly related to the intended function of the horse.

The use of personality assessments must be tested in real-world situations in order to demonstrate their true potential and efficacy. Christensen *et al.* (in press) recently investigated the reliability of a dog temperament test mainly used to identify aggressive dogs. Adopters' of rescued dogs were interviewed up to 13 months post adoption, with regards to their dog's general behaviour with a particular focus on aggressive behaviours. All dogs that were re-homed had passed the temperament test, and yet 40.9% were

reported to have exhibited lunging, growling, snapping and/or biting after adoption. This increased to 71.2% when barking was included in the analysis. The temperament test used to assess the dogs was unable to determine if dogs would show aggressive behaviours in their new homes. These results highlight the importance of quantifying the relevance of temperament and personality tests with respect to real-world situations and should be borne in mind during further development of the HPQ.

It was indicated in Chapter 4 that personality in horses was heritable. There is now increasing evidence to support the genetic inheritance of personality (e.g. Plomin *et al.*, 1994; Loehlin *et al.*, 1998; Bouchard & Loehlin, 2001; Dingemanse *et al.*, 2002; Reif & Lesch, 2003; Stamps, 2003; Sih *et al.*, 2004b; van Oers *et al.*, 2004; Dingemanse & Réale, 2005) but the environmental effects on the selection of behaviour/personality types should not be ignored. The environment that an individual develops in will have some effect on its resulting personality (Suomi, 1987; Archer *et al.*, 2003; Stamps, 2003; Sih *et al.*, 2004b) in combination with its genetic programming. Nevertheless, few studies have focused on these potential factors. Future horse personality research could utilise the HPQ to explore both genetic and environmental effects on horse personality. For example, it was proposed in Chapter 4 that the *Protection* factor may be controlled, to some extent, by maternal influences. Such influences have been discussed by Stamps (2003) who highlighted the current lack of research investigating this area of personality. A deeper understanding of the environmental effects on *Protection* in horses may be of use in the breeding of horses. For instance, do mares that score high on *Protection* make better mothers, and does this in turn affect the care that their daughters provide to their own offspring? Such links have been demonstrated in rodents, such that the maternal behaviour of a female is similar to that of the female that raised her, even if she was not her biological mother (Meaney, 2001; Fleming *et al.*, 2002). If this is the same in horses, then it may explain why the

thoroughbreds and Welsh ponies and cobs became separated from the Arabs on this component alone (See Chapter 4). If this is so, then it may also have important implications for evolutionary concepts of personality.

6.6 Conclusion

In conclusion, the results of this study satisfy all three of Gosling and Vazire's (2002) criteria. As such they not only demonstrate that horse personality can be assessed reliably and with validity, but they have generated several questions with respect to general animal personality research and the potential applications of horse personality assessment. Moreover, the identification of personality differences between breeds provides additional evidence for the heritability of personality. Furthermore significant links between personality and behaviour have been identified and demonstrate the potential for personality assessment to be used to predict behaviour and performance in horses.

Research in animal personality has greatly increased in volume. During the course of this study there has been a noticeable increase in interest in the study of individual differences in behaviour and the reasons for this variation is rapidly being explored. The findings of this study add to this developing knowledge and it is hoped that it will inspire more research in animal and, more specifically, horse personality. With regards to horse personality, future research should focus on developing the HPQ for application within the equine industry.

References

Allport, G. W. & Odbert, H. S. 1936. *Trait-names: A psycho-lexical study*. Psychological Review Company. Princeton, N.J.

Anastasi, A., 1988. Psychological testing (6th Ed.). Macmillan, New York.

Anderson, M. K., Friend, T. H., Evans, J. W. & Bushong, D. M. 1999. Behavioral assessment of horses in therapeutic riding programs. *Applied Animal Behaviour Science*, **63,** 11-24.

Archer, G. S., Friend, T. H., Peidrahita, J., Nevill, C. H. & Walker, S. 2003. Behavioral variation among cloned pigs. *Applied Animal Behaviour Science*, **81,** 321-331.

Arnold, G. W. & Grassia, A. 1952. Ethogram of agonistic behaviour for thoroughbred horses. *Applied Animal Ethology*, **8,** 5-25.

Boguszewski, P. & Zagrodzka, J. 2002. Emotional changes related to age in rats - a behavioral analysis. *Behavioural Brain Research*, **133,** 323-332.

Boissy, A. 1995. Fear and fearfulness in animals. *Quarterly Review of Biology*, **70,** 165-191.

Boissy, A. & Bouissou, M.-F. 1995. Assessment of individual differences in behavioural reactions of heifers exposed to various fear-eliciting situations. *Applied Animal Behaviour Science*, **46,** 17-31.

Borgatta, E. F. 1964. The structure of personality characteristics. *Behavioural Science*, **9,** 8-17.

Both, C., Dingemanse, N. J., Drent, P. J. & Tinbergen, J. M. 2005. Pairs of extreme avian personalities have highest reproductive success. *Journal of Animal Ecology*, **74,** 667-674.

Bouchard, T. J. & Loehlin, J. C. 2001. Genes, Evolution, and personality. *Behavior Genetics*, **31,** 243-273.

Bowling, A. T. & Ruvinsky, A. 2000. Genetic aspects of domestication, breeds and their origins. In: *The genetics of the horse* (Ed. by Bowling, A. T. & Ruvinski, A.), CABI. Oxon, UK, pp. 25-51.

Brace, N., Kemp, R. & Snelgar, R. 2003. *SPSS for psychologists: A guide to data analysis using SPSS for windows*. Palgrave Macmillan.

Bradshaw, J. W. S., Goodwin, D., Lea, A. M. & Whitehead, S. L. 1996. A survey of behavioural characteristics of pure-bred dogs in the United Kingdom. *The Veterinary Record*, **138,** 465-468.

Buckley, P., Dunn, T. & More, S. J. 2004. Owners' perceptions of the health and performance of Pony Club horses in Australia. *Preventative Veterinary Medicine*, **63,** 121-133.

Budaev, S. V. 1999. Sex differences in the Big Five personality factors: Testing an evolutionary hypothesis. *Personality and Individual Differences*, **26,** 801-813.

Buirski, P., Kellerman, H., Plutchik, R., Weininger, R. & Buirski, N. 1973. A field study of emotions, dominance and social behavior in a group of baboons (*Papio anubis*). *Primates*, **14,** 67-78.

Buirski, P. & Plutchik, R. 1991. Measurement of deviant behavior in a Gombe chimpanzee: relation to later behavior. *Primates*, **32,** 207-211.

Buirski, P., Plutchik, R. & Kellerman, H. 1978. Sex differences, dominance and personality in the chimpanzee. *Animal Behaviour*, **26,** 123-129.

Buss, D. M. 1991. Evolutionary personality psychology. *Annual Review of Psychology*, **42,** 459-491.

Byrne, G. & Suomi, S. J. 2002. Cortisol reactivity and its relation to homecage behavior and personality ratings in tufted capuchin (*Cebus apella*) juveniles from birth to six years of age. *Psychoneuroendocrinology*, **27,** 139-154.

Caine, N. G., Earle, H. & Reite, M. 1983. Personality traits of adolescent pig-tailed monkeys (*Macaca nemestrina*): An analysis of social rank and early separation experience. *American Journal of Primatology*, **4,** 253-260.

Capitanio, J. P. 1999. Personality dimensions in adult male rhesus macaques: Prediction of behaviors across time and situation. *American Journal of Primatology*, **47,** 299-320.

Capitanio, J. P., Mendoza, S. P. & Baroncelli, S. 1999. The relationship of personality dimensions in adult male rhesus macaques to progression of Simian Immunodeficiency Virus disease. *Brain, Behavior, and immunity*, **13,** 138-154.

Capitanio, J. P., Mendoza, S. P. & Bentson, K. L. 2004. Personality characteristics and basal cortisol concentrations in adult male rhesus macaques (*Macaca mulatta*). *Psychoneuroendocrinology*, **29,** 1300-1308.

Carere, C., Drent, P. J., Privatera, L., Koohaas, J. M. & Groothuis, T. G. G. 2005. Personalities in Great tits, *Parus major*: stability and consistency. *Animal Behaviour*, **70,** 795-805.

Cattell, R. B. 1943. The description of personality: Basic traits resolved into clusters. *Journal of Abnormal and Social Psychology*, **38,** 476-506.

Cattell, R. B. 1945. The description of personality: Principles and findings in a factor analysis. *American Journal of Psychology*, **58,** 69-90.

Cattell, R. B. 1947. Confirmation and clarification of primary personality factors. *Psychometrika*, **12,** 197-220.

Cattell, R. B. 1973. The isolation of temperament dimensions in dogs. *Behavioral Biology*, **9,** 15-30.

Cavigelli, S. A. 2005. Animal personality and health. *Behaviour*, **142,** 1223-1244.

Chamove, A. S., Eysenck, H. J. & Harlow, H. F. 1972. Personality in monkeys: Factor analysis of rhesus social behaviour. *Quarterly Journal of Experimental Psychology*, **24,** 496-504.

Chaya, L., Cowan, E. & McGuire, B. 2006. A note on the relationship between time spent in turnout and behaviour during turnout in horses (*Equus caballus*). *Applied Animal Behaviour Science*, **98,** 155-160.

Christensen, E. L., Scarlett, J., Campagna, M. & Houpt, K. A. in press. Aggressive behavior in adopted dogs that passed a temperament test. *Applied Animal Behaviour Science*. doi:10.1016/j.applanim.2006.07.002

Christensen, J. W., Ladewig, J., Søndegaard, E. & Malmkavist, J. 2002. Effects of individual versus group stabling on social behaviour in domestic stallions. *Applied Animal Behaviour Science*, **75,** 233-248.

Clarke, S. & Boinski, S. 1995. Temperament in nonhuman primates. *American Journal of Primatology*, **37,** 103-125.

Clutton-Brock, J. 1999. *A natural history of domesticated mammals*. Cambridge University Press. Cambridge, UK

Coolican, H. 2004. *Research methods and statistics in psychology*. Hodder and Stoughton.

Cooper, J. J., McDonald, L. & Mills, D. S. 2000. The effects of increasing visual horizons on stereotypic weaving: implications for social housing of stabled horses. *Applied Animal Behaviour Science*, **69,** 67-83.

Costa, Jr., P. T. & McCrae, R. R. 1992. *Revised NEO personality inventory (NEO-PI-R) and NEO five factor inventory professional manual*. Psychological Assessment Resources. Odessa, FL.

Costa, Jr.,P. T. & McCrae, R. R. 1992. Four way five factors are basic. *Personality and Individual Differences*, **13,** 653-665.

Costa, Jr.,P. T., Terracciano, A. & McCrae, R. R. 2001. Gender differences in personality traits across cultures: robust and surprising findings. *Journal of Personality and Social Psychology*, **81,** 322-331.

Creighton, E. 2003. Matching horses for courses: Development of robust tests of equine temperament to address equine welfare. In: *37th International Congress of The International Society of Animal Ethology*, pp. 148. Venice: Fondazione Iniziative Zooprofilattiche e Zootecniche, Brescia.

Dall, S. R. 2004. Behavioural biology: Fortune favours bold and shy personalities. *Current Biology*, **14,** R470-R472.

Dall, S. R., Houston, A. I. & McNamara, J. M. 2004. The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecology Letters*, **7,** 734-739.

Dean, M. A., Conte, J. M. & Blankenhorn, T. R. in press. Examination of the predictive validity of Big Five Personality dimensions across training performance criteria. *Personality and Individual Differences*.

Diederich, C. & Giffroy, J.-M. 2006. Behavioural testing in dogs: A review of methodology in search for standardisation. *Applied Animal Behaviour Science*, **97,** 51-72.

Digman, J. M. & Takemoto-Chock, N. K. 1981. Factors in the natural language of personality: re-analysis, comparison, and interprtetation of six major studies. *Multivariate Behavioural Research*, **116,** 149-170.

Dingemanse, N. J., Both, C., Drent, P. J., van Oers, K. & van Noordwijk, A. J. 2002. Repeatability and heritability of exploratory behaviour in great tits from the wild. *Animal Behaviour*, **64,** 929-938.

Dingemanse, N.J., Both, C., van Noordwijk, A.J., Rutten, A.L., Drent, P.J., 2003. Natal dispersal and personalities in great tits (*Parus major*). *Proceedings of the Royal Society of London B Biological Sciences*, **270**, 741-747

Dingemanse, N.J., Both, C., Drent., P.J., Tinbergen, J.M., 2004.Fitness consequences of avian personalities in a fluctuating environment. *Proceedings of the Royal Society of London B Biological Science,* **271,** 847-852.

Dingemanse, N. J. & Réale, D. 2005. Natural selection and animal personality. *Behaviour*, **142,** 1159-1184.

Draper, J. 2001. *The ultimate encyclopedia of horse breeds and horse care*. Anness Publishing Ltd. London, UK.

Drent, P. J., van Oers, K. & van Noordwijk, A. J. 2002. Realized heritability of personalities in the great tit (*Parus major*). *Proceedings of the Royal Society of London B Biological Sciences*, **270,** 45-51.

Dutton, D. M., Clark, R. A. & Dickins, D. W. 1997. Personality in captive chimpanzees: use of a novel rating procedure. *International Journal of Primatology*, **18,** 539-552.

Dytham, C. 2003. *Choosing and using statistics: A biologist's guide*. Blackwell Science, Oxford, UK.

Egloff, B. & Gruhn, J. A. 1996. Personality and endurance sports. *Personality and Individual Differences*, **21,** 223-229.

Epstein, S. 1979. The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, **37,** 1097-1126.

Epstein, S. 1983. Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality*, **51,** 360-392.

Eysenck, H.J., 1952. *The scientific study of personality*. Routledge, London, UK

Eysenck, H. J. 1991. Dimensions of personality: 16, 5 or 3? - Criteria for a taxonomic paradigm. *Personality and Individual Differences*, **12,** 773-790.

Eysenck, H.J., 1992. Four ways five factors are not basic. *Personality and Individuals Differences*, **13**. 667-673.

Fairbanks, L. A., Newman, T. K., Bailey, J. N., Jorgensen, M. J., Breidenthal, S. A., Ophoff, R. A., Comuzzie, A. G., Martin, L. J. & Rogers, J. 2004. Genetic contributions to social impulsivity and aggressiveness in vervet monkeys. *Biological Psychiatry*, **55,** 642-647.

Feaver, J., Mendl, M. & Bateson, P. 1986. A method for rating the individual distinctiveness of domestic cats. *Animal Behaviour*, **34,** 1016-1025.

Feh, C. 1988. Social behaviour and relationships of Przewalski horses in Dutch semi-reserves. *Applied Animal Behaviour Science*, **21,** 71-87.

Feh, C. 2005. Relationships and communication in socially natural horse herds. In: (Eds) 2005. *The domestic horse: The evolution, development and management of its behaviour* (Ed. Mills, D., & McDonnell, S.). Cambridge University Press. Cambridge,83-93.

Figueredo, A. J., Cox, R. L. & Rhine, R. 1995. A generalizability analysis of subjective personality assessments in the stumptail macaque and the zebra finch. *Multivariate Behavioural Research*, **30,** 167-197.

Fiske, D. W. 1949. Consistency of the factorial structures of personality from different sources. *Journal of Abnormal and Social Psychology*, **44,** 329-344.

Fleming, A. S., Kraemer, G. W., Gonzalez, A., Lovic, V., Rees, S. & Melo, A. 2002. Mothering begets mothering: The transmission of behaviour and its neurobiology across generations. *Pharmacology Biochemistry and Behavior*, **73,** 61-75.

Foster, V. 2005. *The complete guide to horse breeds: A comprehensive visual directory of horse breeds and horse care*. London, UK: Quantum Publishing.

Fraser, A. F. & Broom, D. M. 1997. *Farm animal behaviour and welfare*. CABI. Oxon, UK.

French, J. M. 1993. Assessment of donkey temperament and the influence of home environment. *Applied Animal Behaviour Science*, **36,** 249-257.

Freud, S. 1961. The ego and the id. In: *The standard edition of the complete psychological works of Sigmund Freud* (Ed. by Strachey, J.), Hogarth Press. London, pp. 1-66.

Funder, D. C. 1995. On the accuracy of personality judgment: a realistic approach. *Psychological Review*, **102,** 652-670.

Funder, D. C. 1999. *Personality judgement: A realistic approach to personality assessment*. Academic Press. San Diego.

Funder, D. C. & Dobroth, K., M. 1987. Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology*, **52,** 409-418.

Funder, D. C., Kolar, D. C. & Blackman, M. C. 1995. Agreement among judges of personality: interpersonal relations, similarity, and acquaintanceship. *Journal of Personality and Social Psychology*, **69,** 656-672.

Gana, K., Trouillet, R., 2003. Structure invariance of the Temperament and Character Inventory. Personality and Individual Differences, 35, 1483-1495.

Garssen, B. 2004. Psychological factors and cancer development: Evidence after 30 years of research. *Clinical Psychology Review*, **24,** 315-338.

Gauly, M., Mathiak, H., Hoffmann, K., Kraus, M. & Erhardt, G. 2001. Estimating genetic variability in temperamental traits in German Angus and Simmental cattle. *Applied Animal Behaviour Science*, **74,** 109-119.

Gillespie, N.A., Cloninger, C.R., Heath, A.C., Martin, N.G., 2003. The genetic and environmental relationship between Cloninger's dimensions of temperament and character. Personality and Individual Differences, 35, 1931-1946

Goddard, M. E. & Beilharz, R. G. 1983. Genetics of traits which determine the suitability of dogs as guide-dogs for the blind. *Applied Animal Ethology*, **9,** 229-315.

Goddard, M. E. & Beilharz, R. G. 1984. A factor analysis of fearfulness in potential guide dogs. *Applied Animal Behaviour Science*, **12,** 253-265.

Goddard, M. E. & Beilharz, R. G. 1986. Early prediction of adult behaviour in potential guide dogs. *Applied Animal Behaviour Science*, **15,** 247-260.

Gold, K. C. & Maple, T. L. 1994. Personality assessment in the gorilla and its utility as a management tool. *Zoo Biology*, **13,** 509-522.

Goldberg, L. R. 1990. An alternative "description of personality": The Big-Five factor structure. *Personality Processes and Individual Differences*, **59,** 1216-1229.

Gosling, S. D. 1998. Personality dimensions in spotted hyenas (*Crocuta crocuta*). *Journal of Comparative Psychology*, **112,** 107-118.

Gosling, S. D. 2001. From mice to men: What can we learn about personality from animal research? *Psychological Bulletin*, **127,** 45-86.

Gosling , S., & Bonnenburg, A.V. 1998. An integrative approach to personality research in anthrozoology: ratings of six species of pets and their owners. *Anthrozoos*, **11**,148-156.

Gosling, S. D. & John, O. P. 1999. Personality dimensions in nonhuman animals: A cross-species review. *Current Directions in Psychological Science*, **8,** 69-75.

Gosling, S. D., John, O. P., Robins, R., W & Craik, K. H. 1998. Do people know how they behave? Self reported act frequencies compared with online codings by observers. *Journal of Personality and Social Psychology*, **74,** 1337-1349.

Gosling, S. D., Kwan, V. S. Y. & John, O. P. 2003. A dog's got personality: A cross-species comparative approach to personality judgments in dogs and humans. *Journal of Personality and Social Psychology*, **85,** 1161-1169.

Gosling, S. D. & Vazire, S. 2002. Are we barking up the right tree? Evaluating a comparative approach to personality. *Journal of Research in Personality*, **36,** 607-614.

Grandin, T., Deesing, M. J., Struthers, J. J. & Swinker, A. M. 1995. Cattle with hair whorl patterns above the eyes are more behaviorally agitated during restraint. *Applied Animal Behaviour Science*, **46,** 117-123.

Gross, R., McIlveen, R., Coolican, H., Clamp, A. & Russel, J. 2000. *Psychology: A new introduction.* Hodder and Stoughton. London.

Hausberger, M., Bruderer, C., Le Scolan, N. & Pierre, J.-S. 2004. Interplay between environmental and genetic factors in temperament/personality traits in horses (*Equus caballus*). *Journal of Comparative Psychology*, **118,** 434-446.

Hausberger, M. & Muller, C. 2002. A brief note on some possible factors involved in the reactions of horses to humans. *Applied Animal Behaviour Science*, **76,** 339-344.

Hayes, K. E. N. 1998. Temperament tip-offs. In: *Horse and Rider*, pp. 46-51.

Hebb, D. O. 1946. Emotion in man and animal: An analysis of the intuitive processes of recognition. *Psychological Review*, **53,** 88-106.

Heird, J. C., Whitaker, D. D., Bell, R. W., Ramsey, C. B. & Lokey, C. E. 1986. The effects of handling at different ages on the subsequent learning ability of 2-year-old horses. *Applied Animal Behaviour Science*, **15,** 15-25.

Heleski, C. R., Shelle, A.C., Nielson, B. D., Zanella, A. J. 2002. Influence of housing on weanling horse behavior and subsequent welfare. *Applied Animal Behaviour Science*, **78,** 291-302.

Henderson, N. D. 1986. Predicting relationships between psychological constructs and genetic characters: An analysis of changing genetic influences on activity in mice. *Behavior Genetics*, **16,** 201-220.

Highland Pony Society. 2006. Highland ponies - Characteristics. www.highlandponysociety.com/highlandponies.htm, accessed 10th July 2006.

Hislop, J. 1992. *Breeding for racing*. London, UK: The Kingswood Press.

Houpt, K. A. & Kusunose, R. 2000. Genetics of behaviour. In: *The genetics of the horse* (Ed. by Bowling, A. T. & Ruvinsky, A.), pp. 281-306. CABI. Oxon, UK.

Irish Draught Horse Society. 2006. Breed Standard. www.irishdraughthorsesociety.com/breed_standards.htm, accessed 10th July 2006

Iguchi, K. I., Matsubara, N. & Hakoyama, H. 2001. Behavioural individuality assessed from two strains of cloned fish. *Animal Behaviour*, **61**, 351-356.

John, O. P. 1990. The 'Big Five' factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In: *Handbook of personality theory and research* (Ed. by Pervin, L. A.), pp. 66-100. New York: Guilford Press.

Jones, A. C. & Gosling, S. D. 2005. Temperament and personality in dogs (*Canis familiaris*): A review and evaluation of past research. *Applied Animal Behaviour Science*, **95,** 1-53.

Kellerman, H. 1966. The emotional behavior of dolphins, *Tursiops truncatus*: Implications for psychoanalysis. *International Mental Health Research Newsletter*, **8,** 2-7.

Kellerman, H. & Plutchik, R. 1968. Emotion-trait interrelations and the measurement of personality. *Psychological Reports*, **23,** 1107-1114.

Kelly, G.A. 1955. *The psychology of personal constructs*. Norton, New York,

Kennedy, J. S. 1992. *The new anthropomorphism.* Cambridge University Press. Cambridge

Kenrick, D. T. & Funder, D. C. 1988. Profiting from controversy: Lessons from the person-situation debate. *American Psychologist*, **43,** 23-34.

Kilgour, R. 1975. The open-field test as an assessment of the temperament of dairy cows. *Animal Behaviour*, **23,** 615-624.

Kilgour, R. J., Melville, G. J. & Greenwood, P. L. in press. Individual differences in the reaction of beef cattle to situations involving social isolation, close proximity of humans, restraint and novelty. *Applied Animal Behaviour Science.*

King, J. E. & Figueredo, A. J. 1997. The five-factor model plus dominance in chimpanzee personality. *Journal of Research in Personality*, **31,** 257-271.

Kline, P. 1993a. *An easy guide to factor analysis*. Routledge. Oxford, UK,

Kline, P., 1993b. *The handbook of psychological testing.* Routledge, London.

Kuhar, C. W., Stoinkski, T. S., Lukas, K., E & Maple, T. L. 2006. Gorilla Behaviour Index revisited: Age, housing and behavior. *Applied Animal Behaviour Science*, **96,** 315-326.

Kummer, H., Dasser, V. & Hoyningen-Huene, P. 1990. Exploring primate social cognition: some critical remarks. *Behaviour*, **112,** 84-98.

Le Scolan, N., Hausberger, M. & Wolff, A. 1997. Stability over situations in temperamental traits of horses as revealed by experimental and scoring approaches. *Behavioural Processes*, **41,** 257-266.

Ley, J., Bennett, P., Coleman, G., in press. Personality dimensions that emerge in companion canines. *Applied Animal behaviour Science*.(doi: 10.1016/j.applanim 2007.04.016)

Lillienfield, S. O., Gershon, J., Duke, M., Marino, L. & de Waal, F. B. M. 1999. A preliminary investigation of the construct of psychopathic personality (psychopathy) in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, **113,** 365-375.

Line, S.W., Hart, B.L., Sanders, L., 1985. Effect of prepubertal versus postpubertal castration on sexual and aggressive behaviour in male horses. *Journal of the American Medical Association*, **186**, (3) 249-251

Loehlin, J. C., McCrae, R. R., Costa Jr, P. T. & John, O. P. 1998. Heritabilities of common and measure-specific components of the Big-Five personality factors. *Journal of Research in Personality*, **32,** 431-453.

Luescher, U. A., McKeown, D. B. & Dean, H. 1998. A cross-sectional study on compulsive behaviour (stable vices) in horses. *Equine veterinary Journal Supplement*, **27,** 14-18.

Mader, D. R. & Price, E. O. 1980. Discrimination learning in horses: Effects of breed, age and social dominance. *Journal of Animal Science*, **50,** 962-965.

Maejima, M., Inoue-Murayama, M., Tonosaki, K., Matsuura, N., Kato, S., Saito, Y., Weiss, A., Murayama, Y., Ito, S., in press. Traits and genotypes may predict the successful training of drug detection dogs. *Applied Animal Behaviour Science,* doi:10.1016/j.applanim.2006.10.005

Maestripieri, D. 2000. Measuring temperament in rhesus macaques: consistency and change in emotionality over time. *Behavioural Processes*, **49,** 167-171.

Malmkvist, J. & Hansen, S. W. 2002. Generalization of fear in farm mink, *Mustela vison*, genetically selected for behavior towards humans. *Animal Behaviour*, **64,** 487-501.

Manteca, X., Deag, J.M., 1993. Social roles in cattle: A plea for interchange of ideas between primatologists and applied ethologists. *Animal Welfare*, **2**, 339-346

Marchetti, C. & Drent, P. J. 2000. Individual differences in the use of social information in foraging by captive great tits. *Animal Behaviour*, **60,** 131-140.

Martau, P. A., Caine, N. G. & Candland, D. K. 1985. Reliability of the Emotions Profile Index, Primate Form, with *Papio hamadryas*, *Macaca fuscata*, and two *Saimiri* species. *Primates*, **26,** 501-505.

Martin, J. E. 2005. The influence of rearing on personality ratings of captive chimpanzees (*Pan troglodytes*). *Applied Animal Behaviour Science*, **90,** 167-181.

Martin, P. & Bateson, P. 1993. *Measuring behaviour: An introductory guide*. Cambridge University Press. Cambridge.

Marušič, A. & Eysenck, H. J. 2001. Synergistic interaction of smoking and neuroticism as a risk factor in ischaemic heart disease: case control study. *Personality and Individual Differences*, **30,** 47-57.

Mather, J. A. & Anderson, R. C. 1993. Personalities of octopuses (*Octopus rubescens*). *Journal of Comparative Psychology*, **107,** 336-340.

McAdam, A. G., Boutin, S., Réale, D. & Berteaux, D. 2002. Maternal effects and the potential for evolution in a natural population of animals. *Evolution*, **56,** 846-851.

McAfee, L. M., Mills, D. S. & Cooper, J. J. 2002. The use of mirrors for the control of stereotypic weaving behaviour in the stabled horse. *Applied Animal Behaviour Science*, **78,** 159-173.

McBride, S. D. & Wolf, B. in press. Using multivariate statistical analysis to measure ovine temperament; stability of factor contruction over time and between groups of animals. *Applied Animal Behaviour Science.*

McCall, C. A., Hall, S., McElhenney, W. H. & Cummin, K. A. 2006. Evaluation and comparison of four methods of ranking horses based on reactivity. *Applied Animal Behaviour Science*, **96,** 115-127.

McCann, J. S., Heird, J. C., Bell, R. W. & Lutherer, L. O. 1988. Normal and more highly reactive horses. I. Heart rate, respiration rate and behavioral observations. *Applied Animal Behaviour Science*, **19,** 201-214.

McCrae, R. R. 2002. The maturation of personality psychology: Adult personality development and psychological well-being. *Journal of Research in Personality*, **36** 307-317.

McCrae, R. R. & Costa, Jr., P. T., 1985. Comparison of EPI and psychoticism scales with measures of the five-factor model of personality. *Personality and Individual Differences*, **6,** 587-597.

McCrae, R. R. & Costa, Jr., P. T., 1987. Validation of the Five-Factor Model of Personality across instruments and observers. *Journal of Personality and Social Psychology*, **52,** 81-90.

McCrae, R. R. & Costa, Jr., P. T., 1999. A five-factor theory of personality In: *Handbook of personality: Theory and research* (Ed. by Pervin, L. A. & John, O. P.), Guilford Press. New York, pp. 139-153.

McCrae, R. R. & Costa, Jr., P. T., 2004. A contemplated revision of the NEO Five Factor Inventory. *Personality and Individual Differences*, **36,** 587-596.

McDonnell, S. M., 2005. Sexual Behaviour. In: *The domestic horse: The evolution, development and management of its behaviour* (Ed. Mills, D., McDonnell, S.,). Cambridge University Press, Cambridge.

McDonnell, S. M. 2002. Behaviour of horses. In: *The ethology of domestic animals: An introductory text* (Ed. by Jensen, P.), pp. 119-129. CABI publishing. Oxon, UK.

McDonnell, S. M. & Haviland, J. C. S. 1995. Agonistic ethogram of the equid bachelor band. *Applied Animal Behaviour Science*, **43,** 147-188.

McDonnell, S. M. & Poulin, A. 2002. Equid play ethogram. *Applied Animal Behaviour Science*, **78,** 263-290.

McDougall, P. T., Réale, D., Sol, D. & Reader, S. 2006. Wildlife conservation and animal temperament: causes and consequences of evolutionary change for captive, reintroduced, and wild populations. *Animal Conservation*, **9,** 39-48.

McElreath, R., Strimling, P., 2006. How noisy information and individual asymmetries can make 'personality' an adaptation: a simple model. *Animal Behaviour*, **72,** 1135-1139.

McGreevy, P., 2004. *Equine behaviour: A guide for veterinarians and equine scientists*. Saunders, London.

McGreevy, P. & Nicol, C. 1998. Physiological and behavioral consequences associated with short-term prevention of crib-biting in horses. *Physiology and Behavior*, **65,** 15-23.

McGreevy, P. D. & Thomson, P. C. 2006. Differences in motor laterality between breeds of performance horse. *Applied Animal Behaviour Science.*

McGrogan, C., Hutchinson, M.D., King, J.E. in press. Dimensions of horse personality based on owner and trainer supplied personality traits. *Applied Animal Behaviour Science*, (doi:10.1016/j.applanim.2007.10.006)

McGuire, M. T., Raleigh, M. J. & Pollack, D. B. 1994. Personality features in vervet monkeys: The effects of sex, age, social status and group composition. *American Journal of Primatology*, **33,** 1-13.

Meaney, M. J. 2001. Maternal care, gene expression and the transmission of individual differences in stress reactivity across generations. *Annual Review of Neuroscience*, **24,** 1161-1192.

Mettke-Hofmann, C., Ebert, C., Schmidt, T., Steiger, S. & Stieb, S. 2005. Personality traits in resident and migratory warbler species. *Behaviour*, **142**.

Mills, D. S. 1998. Personality and individual differences in the horse, their significance, use and measurement. *Equine Veterinary Journal Supplement*, **27,** 10-13.

Mischel, W. 1968. *Personality and assessment*. Wiley. New York.

Mitchell, R. W. & Hamm, M. 1997. The interpretation of animal psychology: anthropomorphism or behavior reading? *Behaviour*, **134,** 173-204.

Momozawa, Y., Kusunose, R., Kikusui, T., Takeuchi, Y. & Mori, Y. 2005. Assessment of equine temperament questionnaire by comparing factor structure between two separate surveys. *Applied Animal Behaviour Science*, **92,** 77-84.

Momozawa, Y., Ono, T., Sato, F., Kikusui, T., Takeuchi, Y., Mori, Y. & Kusunose, R. 2003. Assessment of equine temperament by a questionnaire survey to caretakers and evaluation of its reliability by simultaneous behavior test. *Applied Animal Behaviour Science*, **84,** 127-138.

Mormède, P. 2005. Molecular genetics of behaviour: research strategies and perspectives for animal production. *Livestock production science*, **93,** 15-21.

Morris, P. H., Gale, A. & Duffy, K. 2002a. Can judges agree on the personality of horses? *Personality and Individual Differences*, **33,** 67-81.

Morris, P.H., Gale, A., Howe, S., 2002b. The factor structure of horse personality. Anthrozoös, 15 (4), 300-322.

Moutafi, J., Furnham, A. & Paltiel, L. 2005. Can personality factors predict intelligence? *Personality and Individual Differences*, **38,** 1021-1033.

Muller, R. & Schrader, L. 2005. Behavioural consistency during social separation and personality in dairy cows. *Behaviour*, **142,** 1289-1306.

Murray, L. E. 1998. The effects of group structure and rearing strategy on personality in chimpanzees (Pan troglodytes) at Chester, London ZSL and Twycross zoos. *International Zoo Yearbook*, **36,** 97-108.

Ninomiya, S., Sato, S., Kusunose, R., Mitumasu, T. & Obara, Y. in press. A note on a behavioural indicator of satisfaction in stabled horses. *Applied Animal Behaviour Science.*

Norman, W. T. 1963. Toward an adequate taxonomy of personality attributes: replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, **66,** 574-583.

Norman, W. T. 1967. 2800 personality trait descriptors: Normative operating characteristics for a university population. In: *Department of Psychology*: University of Michigan.

Notari, L. & Goodwin, D. in press. A survey of behavioural characteristics of pure-bred dogs in Italy. *Applied Animal Behaviour Science.*

Pederson, A. K., King, J. E. & Landau, V. I. 2005. Chimpanzee (*Pan troglodytes*) personality predicts behaviour. *Journal of Research in Personality*, **39,** 534-549.

Pérez-Guisado, J., Lopex-Rodríguez, R. & Muñoz-Serrano, A. 2006. Heretibiltiy of dominant-aggressive behaviour in English Cocker-Spaniels. *Applied Animal Behaviour Science*, **100,** 219-227.

Pervin, L.A., Cervone, D., John, O.P., 2005. *Personality theory and research (9th Ed.).* John Wiley and Sons, New York

Pervin, L. A. & John, O. P. 1997. *Personality theory and research.* John Wiley and Sons. New York.

Platman, J. R., Plutchik, R., Fieve, R. R. & Lawlor, W. G. 1969. Emotion profiles associated with mania and depression. *Archives of General Psychiatry*, **20,** 210-214.

Plomin, R., Owen, M. J. & McGuffin, P. 1994. The genetic basis of complex human behaviors. *Science*, **264**, 1733-1739.

Plutchik, R. 1960. The multifactor-analytic theory of emotion. *The Journal of Psychology*, **50,** 153-171.

Plutchik, R. 1965. What is an emotion? *Journal of Psychology*, **61,** 295-303.

Ramsay, M.C., Reynolds, C.R., 2000. Development of a scientific test: A practical guide. In: Goldstein, G., & Hersen, M., (Eds) Handbook of psychological assessment 3rd ed. Elsevier Science Ltd.

Reckase, M.D., 2000. Scaling techniques. In Goldstein, G., & Hersen, M., 2000. *Handbook of psychological assessment 3rd Ed.* Elsevier Science Ltd.

Redbo, I., Redbo-Torstensson, P., Odberg, F. O., Hedendahl, A. & Holm, J. 1998. Factors affecting behavioural disturbances in race-horses. *Animal Science*, **66,** 475-481.

Reif, A. & Lesch, K.-P. 2003. Toward a molecular architecture of personality. *Behavioural Brain Research*, **139,** 1-20.

Rivera, E., Benjamin, S., Nielsen, B., Shelle, J. & Zanella, A. J. 2002. Behavioural and physiological responses of horses to initial training: the comparison between pastured versus stalled horses. *Applied Animal Behaviour Science*, **78,** 235-252.

Romeyer, A. & Bouissou, M.-F. 1992. Assessment of fear reactions in domestic sheep, and influence of breed and rearing conditions. *Applied Animal Behaviour Science*, **34,** 93-119.

Rouff, J. H., Sussman, R. W. & Strube, M. J. 2005. Personality traits in captive lion-tailed macaques (*Macaca silenus*). *American Journal of Primatology*, **67,** 177-198.

Rowe, P. M. 1995. Neural networks: a bridge between neuroscience and psychology. In: *Molecular Medicine Today*, **18**, pp. 167-173.

Ruis, M. A. W., te Brake, J. H. A., van de Burgwal, J. A., de Jong, I. C., Blokhuis, H. J. & Koolhaas, J. M. 2000. Personalities in female domesticated pigs: behavioural and physiological indications. *Applied Animal Behaviour Science*, **66,** 31-47.

Seaman, S. C., Davidson, H. P. B. & Waran, N. K. 2002. How reliable is temperament assessment in the domestic horse (*Equus caballus*)? *Applied Animal Behaviour Science*, **78,** 175-191.

Serpell, J. A. & Hsu, Y. 2001. Development and validation of a novel method for evaluating behavior and temperament in guide dogs. *Applied Animal Behaviour Science*, **72**, 347-364.

Serpell, J. A. & Hsu, Y. 2005. Effects of breed, sex, and neuter status on trainability in dogs. *Anthrozoos*, **18,** 196-207.

Sibley, C. G. & Ahlquist, J. E. 1984. The phylogeny of the hominid primates, as indicated by DNA-DNA hybridization. *Journal of Molecular Evolution*, **20,** 2-15.

Siegel, S. & Castellan, N. 1988. *Nonparametric statistics for the behavioural sciences*. New York: McGraw-Hill.

Sih, A., Bell, A. M., Johnson, J. C. & Ziemba, R. E. 2004a. Behavioural syndromes: An ecological and evolutionary overview. *Trends in Ecology and Evolution*, **19,** 372-378.

Sih, A., Bell, A. M., Johnson, J. C. & Ziemba, R. E. 2004b. Behavioural syndromes: An integrative overview. *The Quarterly Review of Biology*, **79,** 241-277.

Sinn, D. L., Mather, J. A. & Anderson, R. C. 2001. Early temperamental traits in an octopus (*Octopus bimaculoides*). *Journal of Comparative Psychology*, **115,** 351-364.

Spoolder, H. A. M., Burbidge, J. A., Lawrence, A. B., Simmins, P. H. & Edwards, S. A. 1996. Individual behavioural differences in pigs: Intra-and inter-test consistency. *Applied Animal Behaviour Science*, **49,** 185-198.

Srivastava, S., John, O. P., Gosling, S. D. & Potter, J. 2003. Development of personality in early and middle adulthood: Set like a plaster or persistent change. *Journal of Personality and Social Psychology*, **84,** 1041-1053.

Stamps, J. 2003. Behavioural processes affecting development: Tinbergen's fourth question comes of age. *Animal Behaviour*, **66,** 1-13.

Stevenson-Hinde, J., Stillwell-Barnes, R. & Zunz, M. 1980. Subjective assessment of rhesus monkeys over four successive years. *Primates*, **21,** 66-82.

Stevenson-Hinde, J. & Zunz, M. 1978. Subjective assessment of individual rhesus monkeys. *Primates*, **19,** 473- 482.

Strand, S. C., Tiefenbacher, S., Haskell, M., Hosmer, T., McDonnell, S. M. & Freeman, D. A. 2002. Behavior and physiologic responses of mares to short-term isolation. *Applied Animal Behaviour Science*, **78,** 145-157.

Strandberg, E., Jacobsson, J. & Saetre, P. 2005. Direct genetic, maternal and litter effects on behaviour in German shepherd dogs in Sweden. *Livestock production science*, **93,** 33-42.

Suomi, S.J., 1987. Genetic and maternal contributions to individual differences in rhesus monkey biobehavioural development. In: Krasnegar, N., Blass, E., Hofer, M., Smotherman, W. (eds), *Perinatal development : A psychological perspective*. New York, Academic Press. 397-420

Suomi, S. J., Novak, M. A. & Well, A. 1996. Aging in rhesus monkeys: Different windows on behavioural continuity and change. *Developmental Psychology*, **32,** 1116-1128.

Svartberg, K. 2002. Shyness-boldness predicts performance in working dogs. *Applied Animal Behaviour Science*, **79,** 157-174.

Svartberg, K. 2005. A comparison of behaviour in test and in everyday life: Evidence of three consistent boldness-related personality traits in dogs. *Applied Animal Behaviour Science*, **91,** 103-128.

Svartberg, K. 2006. Breed-typical behaviour in dogs - historical remnants or recent constructs. *Applied Animal Behaviour Science*, **96,** 293-313.

Svartberg, K. & Forkman, B. 2002. Personality traits in the domestic dog (*Canis familiaris*). *Applied Animal Behaviour Science*, **79,** 133-155.

Svartberg, K., Tapper, I., Temrin, H., Radesäter, T. & Thornman, S. 2005. Consistency of personality traits in dogs. *Animal Behaviour*, **69,** 283-291.

Tabachnick, B. G. & Fidell, L. S. 2001. *Using multivariate statistics*. New York: HarperCollins.

van Oers, K., De Jong, G., Drent, P. J. & van Noordwijk, A. J. 2004. A genetic analysis of avian personality traits: correlated, response to artificial slection. *Behavior Genetics*, **34,** 611-619.

van Oers, K., de Jong, G., van Noordwijk, A. J., Kempenaers, B. & Drent, P. J. 2005. Contribution of genetics to the study of animal personalities: a review of case studies. *Behaviour*, **142,** 1185-1206.

van Oers, K., Drent, P. J., de Goede, P. & van Noordwijk, A. J. 2003. Realized heritability and repeatability of risk-taking behaviour in relation to avian personalities. *Proceedings of the Royal Society of London*, **271,** 65-73.

Verbeek, M.E.M., Drent, P.J., Wiepkema, P.R., 1994. Consistent individual differences in early exploratory behaviour in male great tits. *Animal Behaviour*, **48,** 1113-1121.

Visser, E. K., van Reenen, C. G., Engel, B., Schilder, M. B. H., Barneveld, A. & Blokhuis, H. J. 2003a. The association between performance in show-jumping and personality traits earlier in life. *Applied Animal Behaviour Science*, **82,** 279-295.

Visser, E. K., van Reenen, C. G., Hopster, H., Schilder, M. B. H., Knaap, J. H., Barneveld, A. & Blokhuis, H. J. 2001. Quantifying aspects of young horses' temperament: consistency of behavioural variables. *Applied Animal Behaviour Science*, **74,** 241-258.

Visser, E. K., van Reenen, C. G., Schilder, M. B. H., Barneveld, A. & Blokhuis, H. J. 2003b. Learning performances in young horses using two different learning tests. *Applied Animal Behaviour Science*, **80,** 311-326.

Visser, E. K., van Reenen, C. G., van der Werf, J. T. N., Schilder, M. B. H., Knaap, J. H., Barneveld, A. & Blokhuis, H. J. 2002. Heart rate and heart rate variability during a novel object test and a handling test in young horses. *Physiology and Behavior*, **76,** 289-296.

Weiss, A., King, J. E. & Enns, R. M. 2002. Subjective well-being is heritable and genetically correlated with dominance in chimpanzees (*Pan troglodytes*). *Personality Processes and Individual Differences*, **83,** 1141-1149.

Weiss, A., King, J. E. & Figueredo, A. J. 2000. The heritability of personality factors in chimpanzees (*Pan troglodytes*). *Behavior Genetics*, **30,** 213-221.

Westen, D. 1990. Psychoanalytic approaches to personality. In: *Handbook of personality theory and research* (Ed. by Pervin, L. A.), pp. 21-65. Guilford Press. New York.

Wielebnowski, N. C. 1999. Behavioural differences as predictors of breeding status in captive cheetahs. *Zoo Biology*, **18,** 335-349.

Wolf, M., van Doorn,G.S., Leimar, O., Weissing, F.J., 2007. Life-history trade-offs favour the evolution of animal personalities.Nature, **447**, 581-585.

Wolff, A. & Hausberger, M. 1996. Learning and memorisation of two different tasks in horses: the effects of age, sex and sire. *Applied Animal Behaviour Science*, **46,** 137-143.

Wolff, A., Hausberger, M. & Le Scolan, N. 1997. Experimental tests to assess emotionality in horses. *Behavioural Processes*, **40,** 209-221.

Zar, J.H., 1999. *Biostatistical analysis 4th Ed.* Prentice Hall, New Jersey.

Zuckerman, M. 1991a. Behavioral genetics and personality traits. In: *Psychobiology of personality* (Ed. by Zuckerman, M.), Cambridge University Press. Cambridge, UK. pp. 89-129.

Zuckerman, M. 1991b. Consistency in personality. In: *Psychobiology of personality* (Ed. by Zuckerman, M.), Cambridge University Press. Cambridge, UK, pp. 44-88.

Appendices

Appendix 1: Rater reliability between horses after the removal of five unreliable traits: Kendall's W and Spearman rank order correlation coefficients (r_s).

Horse	Gender	Age	Breed	Yard	*W*	r_s coefficient			Rated reliably?
						R1 R2	R1 R3	R2 R3	
1	M	27	U	2	0.53*	-	-	-	YES
2	F	13	TBx	2	0.24	-0.44*	-0.43*	0.51*	YES
3	M	8	TB	2	0.61**	-	-	-	YES
4	M	23	TB	2	0.59**	-	-	-	YES
5	M	13	IDx	2	0.67**	-	-		YES
6	M	5	SH	2	0.39	0.11	-0.19	0.40*	YES
7	F	10	TBx	2	0.35	0.18	-0.08	0.13	NO
8	F	10	TB	2	0.43	-0.05	0.45*	0.34	YES
9	M	10	Fres	2	0.62**	-	-	-	YES
10	M	6	SH	2	0.49**	-	-	-	YES
11	M	5	Con x	2	0.65**	-	-	-	YES
12	M	10	TB	2	0.39	0.06	-0.26	0.53**	YES
13	F	13	TBx	2	0.59**	-	-	-	YES
14	F	13	TBx	2	0.36	-0.21	0.03	0.39	NO
15	M	8	TB	2	0.43	0.11	0.01	0.40	NO
16	F	3	TBx	2	0.74***	-	-	-	YES
17	M	17	SH	2	0.54**	-	-	-	YES
18	F	12	AA	2	0.43	0.19	0.16	0.22	NO
19	F	13	Arab	2	0.34	-0.12	0.14	0.17	NO
20	F	12	U	1	0.50*	-0.05	0.57**	0.37	YES
21	F	10	U	1	0.63**	-	-	-	YES
22	F	7	TBx	1	0.47	0.27	0.11	0.39	NO
23	F	15	U	1	0.68***	-	-	-	YES
24	F	10	TBx	1	0.67***	-	-	-	YES
25	F	12	TBx	1	0.46	0.20	0.24	0.25	NO
26	F	12	U	1	0.72***	-	-	-	YES
27	F	14	U	1	0.39	0.06	0.18	0.19	NO
28	F	19	U	1	0.36	-0.10	0.36	-0.02	NO
29	M	19	U	1	0.56*	-	-	-	YES
30	M	11	TBx	1	0.57**	-	-	-	YES

* $P<0.05$, ** $P<0.01$, *** $P<0.001$; r_s only calculated when *W* not significant ($P>0.05$), M = male (gelding); F = female (mare); R1 = rater one; R2 = rater two; R3 = rater three; U = unknown breeding; TB = thoroughbred, TBx = thoroughbred cross; IDx = Irish draught cross; SH = sports horse; Fres = Friesian; Con x = Connemara cross; AA = Anglo Arab; Arab = Arabian horse.

Appendix 1 (continued): Rater reliability between horses, after the removal of the five unreliable traits: Kendall's *W* and Spearman rank order correlation coefficients (r_s)

Horse	Gender	Age	Breed	Yard	*W*	r_s coefficient			Rated reliably?
						R1 R2	R1 R3	R2 R3	
31	M	7	TB	1	0.46	0.01	0.12	0.52**	YES
32	M	6	U	1	0.78***	-	-	-	YES
33	M	14	U	1	0.53**	-	-	-	YES
34	M	7	TB	1	0.61**	-	-	-	YES
35	M	15	U	1	0.69***	-	-	-	YES
36	M	-	U	1	0.30	0.10	-0.31	0.14	NO
37	M	-	U	1	0.61**	-	-	-	YES
38	M	15	U	1	0.53*	-	-	-	YES
39	M	-	U	1	0.36	-0.08	-0.32	0.66**	YES
40	F	16	U	3	0.33	-0.15	0.26	-0.05	NO
41	M	5	U	3	0.57**	-	-	-	YES
42	M	8	TB	3	0.56*	-	-	-	YES
43	M	18	DWB	3	0.61**	-	-	-	YES
44	M	5	TBx	3	0.36	0.23	0.06	-0.02	NO
45	M	8	TB	3	0.42	0.08	0.22	0.25	NO
46	F	5	U	3	0.24	0.04	-0.28	-0.11	NO
47	M	5	U	3	0.44	0.30	0.339	-0.04	NO
48	M	19	TBx	3	0.30	-0.34	0.71**	-0.43*	YES
49	M	6	WC	3	0.61**	-	-	-	YES
50	F	16	DWB	3	0.46	0.18	0.33	0.24	NO
51	M	15	Arab	4	0.34	-0.22	0.028	0.40*	YES
52	M	6	TBx	4	0.55*	-	-	-	YES
53	M	16	TBx	4	0.53*	-	-	-	YES
54	M	17	TB	4	0.25	-0.29	-0.39	0.54*	YES
55	M	4	IDH	4	0.49*	-	-	-	YES
56	F	21	TBx	4	0.65**	-	-	-	YES
57	M	20	TBx	4	0.53*	-	-	-	YES
58	F	10	TB	4	0.59**	-	-	-	YES
59	M	16	TBx	4	0.40	0.15	0.21	0.10	NO
60	M	15	TB	4	0.41	0.42*	-0.12	0.16	YES
61	M	8	DWB	4	0.65**	-	-	-	YES

* $P<0.05$, ** $P<0.01$, *** $P<0.001$; r_s only calculated when *W* not significant ($P>0.05$)

M = male/gelding; F = female/mare; R1 = rater one; R2 = rater two; R3 = rater three; U = unknown breeding; TB = thoroughbred, TBx = thoroughbred cross; IDx = Irish draught cross; SH = sports horse; Arab = Arabian horse; WC = Welsh section C; DWB = Danish warm blood; IDH = Irish draught horse.

Appendix 2: Horse Personality Questionnaire

Horse Personality Questionnaire 2003

By Adele Lloyd

IN ASSOCIATION WITH:

Moulton College

Open University

Harper Adams University College

Instructions for Horse Personality Questionnaire

- Someone who regularly handles the horse (i.e. Minimum of twice a week) and has done so for a minimum of one year should fill in the questionnaire. This may or may not be the owner.

- More than one person may fill in a questionnaire for a particular horse as long as they meet the above requirements and do not discuss their answers.

- Answers must reflect the INDIVIDUAL'S opinion of the horse's personality. PLEASE DO NOT CONFER!

- Try to be as accurate as possible when describing the horse's personality. Think about the question and try to give the horse the 'score' you most think best describes it.

Thank you for your cooperation and time, Adele Lloyd.

ABOUT YOU:

Name:…… Date of birth: ……

Age: …… Gender: ……

How long have you known the horse? ……

Are you the current owner of the horse? If not state your connection:

……

……

ABOUT THE HORSE:

Name: ……

Date of birth: ……Age: ……

Breed: …… Sex:……

Time of current ownership: ……

Time kept at current location: ……

How many people **REGULARLY** handle the horse? ……

How are the other handlers involved with the horse? (Grooming, riding, training etc?).

……

……

……

……

PLEASE ANSWER THE FOLLOWING QUESTIONS USING A SCALE OF 1 TO 7.

- WHERE 1 INFERS NO EXPRESSION AND 7 INFERS EXTREME/TOTAL EXPRESSION.
- A SCORE OF ANY NUMBER BETWEEN 1 AND 7 CAN BE GIVEN BUT CHOOSE THE NUMBER THAT YOU THINK BEST DESCRIBES THE HORSE IN QUESTION
- PLEASE READ EACH QUESTION THOROUGHLY AND CONSIDER YOUR ANSWER CAREFULLY WITHOUT CONFERRING.

Example 1

Is the horse ACTIVE? (Moves around a lot and doesn't like being still for long).

1 = Not active

2 = very rarely active

3 = scarcely active

4 = 50% time active 50% time inactive.

5 = quite active

6 = regularly active

7 = extremely active

Example 2

Is the horse equable? (Very composed and reacts to others, horse and/or human, in an even and calm manner, and is not easily bothered or worried).

1 = not equable

2 = very rarely equable

3 = scarcely equable

4 = 50% time equable.

5 = quite equable

6 = regularly equable

7 = extremely equable

PLEASE ANSWER THE FOLLOWING QUESTIONS REFERRING TO THE HORSE'S PERSONALITY BY CIRCLING THE MOST APPROPRAITE NUMBER

1. Is the horse **ACTIVE**? (Moves around a lot and will not stay still for very long).
   Low 1 2 3 4 5 6 7 **High**

2. Is the horse **AGGRESSIVE**? (Causes or threatens to cause potential harm to other individuals, i.e. humans, horses or other animals).
   Low 1 2 3 4 5 6 7 **High**

3. Is the horse **APPREHENSIVE**? (Appears to be anxious about everything and fears or avoids any kind of risks).
   Low 1 2 3 4 5 6 7 **High**

4. Is the horse **CONFIDENT**? (Behaves in a positive, assured manner, and will not appear restrained or tentative about its actions).
   Low 1 2 3 4 5 6 7 **High**

5. Is the horse **CURIOUS**? (readily explores new situations or objects).
   Low 1 2 3 4 5 6 7 **High**

6. Is the horse **ECCENTRIC**? (Shows stereotypies, unusual mannerisms and exaggerated behaviour).
   Low 1 2 3 4 5 6 7 **High**

7. Is the horse **EFFECTIVE**? (Dominant individual, will regularly get its own way and has an ability to control the behaviour of others).
   Low 1 2 3 4 5 6 7 **High**

8. Is the horse **EQUABLE**? (Very composed and reacts to others, horse and/or human, in an even and calm manner, and is not easily bothered or worried).
   Low 1 2 3 4 5 6 7 **High**

9. Is the horse **EXCITABLE**? (Over reacts to changes, is highly strung and is easily excited or wound up).
   Low 1 2 3 4 5 6 7 **High**

10. Is the horse **FEARFUL**? (Retreats readily from others or outside disturbances and spooks easily).
    Low 1 2 3 4 5 6 7 **High**

11. Is the horse **HARD WORKING**? (Appears keen to do well during work/exercise and concentrates on what it is being asked to do, it may also respond well to instructions).
    Low 1 2 3 4 5 6 7 **High**

12. Is the horse **INSECURE**? (Hesitates to act alone and might seek reassurance from others and tends to be more confident or settled when with other horses).
    Low 1 2 3 4 5 6 7 **High**

13. Is the horse **INTELLIGIENT**? (Learns new things, [i.e. skills, commands etc.] easily and quickly and may have an ability to problem solve. May also appear to benefit from mental stimulation).

Low 1 2 3 4 5 6 7 **High**

14. Is the horse **IRRITABLE**? (Reacts negatively with little provocation, is highly eruptive and volatile).

Low 1 2 3 4 5 6 7 **High**

15. Is the horse **MOTHERLY**? (Provides a warm, receptive and secure base for others, is tender and caring).

Low 1 2 3 4 5 6 7 **High**

16. Is the horse **OPPORTUNISTIC**? (Seizes a chance as soon as it arises. Will take advantage of a situation).

Low 1 2 3 4 5 6 7 **High**

17. Is the horse **PERMISSIVE**? (Could, but does not interfere with behaviour of others).

Low 1 2 3 4 5 6 7 **High**

18. Is the horse **PLAYFUL**? (Initiates play and joins in when play is solicited).

Low 1 2 3 4 5 6 7 **High**

19. Is the horse **POPULAR**? (Others seek it out as a companion).

Low 1 2 3 4 5 6 7 **High**

20. Is the horse **PROTECTIVE**? (Will prevent harm or possible harm to others, defensive of others, human and/or horse).

Low 1 2 3 4 5 6 7 **High**

21. Is the horse **RELIABLE**? (Can be trusted to do things or behaves well might also be considered a safe horse to be around).

Low 1 2 3 4 5 6 7 **High**

22. Is the horse **SLOW**? (Moves and rests in a relaxed manner, moves slowly but deliberately and is not easily hurried).

Low 1 2 3 4 5 6 7 **High**

23. Is the horse **SOCIABLE**? (Seeks the companionship of others and rarely separates from the group by choice).

Low 1 2 3 4 5 6 7 **High**

24. Is the horse **SOLITARY**? (Spends a lot of time alone, by choice and will deliberately separate from the rest of the herd/group).

Low 1 2 3 4 5 6 7 **High**

25. Is the horse **STRONG**? (Relies on and uses its own strength and sturdiness regularly).

Low 1 2 3 4 5 6 7 **High**

26. Is the horse **STUBBORN**? (Does not give in readily or easily, and is not very co-operative with others, human and/or horse).

Low 1 2 3 4 5 6 7 **High**

27. Is the horse **SUBORDINATE**? (Gives in readily to others and submits easily. Will not put up a fight, gets out of the way quickly).

Low 1 2 3 4 5 6 7 **High**

28. Is the horse **SUSPICIOUS** (of others)? (Doesn't trust others readily, human and/or horse, trusts few/select individuals).

Low 1 2 3 4 5 6 7 **High**

29. Is the horse **TENSE**? (Restrained in posture and movement; carries the body stiffly which, suggests a shrinking tendency as if pulling back to be less conspicuous).

Low 1 2 3 4 5 6 7 **High**

30. Is the horse **UNDERSTANDING**? (Responds in a discriminating and appropriate manner to the behaviour of others. Shows a sense of understanding/comprehension/consideration).

Low 1 2 3 4 5 6 7 **High**

Thank you for completing this questionnaire, if you have any problems or would like more information, please contact Adele Lloyd: Tel. 01604 491131 ext. 608 or email adeleL@moulton.ac.uk

Once completed please return the questionnaire to:

Adele Lloyd, Moulton College, Moulton, Northampton, NN3 7RR

Appendix 3: Ethogram of horse behaviours measured during field observations

Behaviour	Definition
Stand	Stand with eyes either down or forward, ears soft for 5s or more. (Strand *et al.*, 2002)
Vigilance	Stand with eyes focused forward for 5s or more (Strand *et al.*, 2002). Elevated neck intently orientated head and ears (Le Scolan *et al.*, 1997). Ears held stiffly upright nostrils dilated (McDonnell & Haviland, 1995)
Walk	Move forward with slow four beat gait. (Strand *et al.*, 2002). Horse walks energetically, looks in front of self and around (Le Scolan *et al.*, 1997)
Trot	Two beat gait (Strand *et al.*, 2002)
Canter	Three beat gait (Strand *et al.*, 2002)
Gallop	Fast four beat gait (Strand *et al.*, 2002)
Passage	Animated form of trot where the legs are raised with more elevation often associated with audible hoof contact with the ground (Le Scolan *et al.*, 1997)
Spook	Move abruptly in any direction in a manner typical of avoidance or removal from an area (Strand *et al.*, 2002)
Exploration	Horse walks slowly with neck horizontally or lower, ready to stop and sniff the ground or other object, characteristic walk of a quiet horse in a calm situation (Le Scolan *et al.*, 1997)
Bitten/biting	Opening and rapid closing of the jaws with the teeth grasping the flesh of another horse, ears are pinned and lips retracted. Includes bite threats used more as a warning (McDonnell & Haviland, 1995). Biting = Individual biting another; Bitten = individual bitten by another
Kicked/kicking	One or more legs lifted off the ground and rapidly extended towards another individual with apparent intent to make contact, also includes kick threat where no contact is made; Kicking = to inflict against another individual; Kicked = To receive a kick or kick threat
Nipped/nipping	Similar to bite, but mouth less widely open and teeth closing on small piece of flesh (McDonnell and Haviland, 1995); Nipped = receive a nip; Nipping = to nip another individual
Chased/chasing	One individual pursues another, with an apparent attempt to overtake or catch up. Chaser typically may show bite threats or bite the chased individual (McDonnell & Haviland, 1995). Chasing = individual chasing another; Chased = individual being chased
Head threat given/head threat received	Head lowered with ears pinned back, neck stretched or extended toward the target individual and often lips pursed (McDonnell & Haviland, 1995); HTG = Head threat given; HTR = Head threat received)

Appendix 3 (continued): Ethogram of horse behaviours measured during field observations

Behaviour	Definition
Herded/herding	Combination of head threat and ears laid back with forward locomotion, apparently directing the movement of another individual or individuals (McDonnell & Haviland, 1995); Herding = Individual herding others; Herded = Individual being herded
Play fight	Involves behavioural elements and sequencing similar to adult fighting behaviour - often involves nipping, biting, and kicking but without serious harm (McDonnell & Poulin, 2002)
Self groom	Individual rubs body against stationary object (i.e. fence post etc.), nip body with teeth or scratch with hoof (Strand *et al.*, 2002).
Mutual groom	Two individuals stand side by side usually head to tail and groom each other. Groom-I = Individual initiates mutual grooming; Groom-R = Individual takes part in mutual grooming but did not initiate session
Roll	Dropping from standing to sternal recumbency, then rotating one or more times from sternal to dorsal recumbence, tucking legs against the body (McDonnell and Haviland, 1995)
Greet	Horses stand face to face and sniff various parts of each other's head and body, often combined with short high pitched vocalisation (McDonnell and Haviland, 1995)
Graze	With head lowered to ground, take grass into mouth and chew (for 5s or longer) (Strand *et al.*, 2002). May also move step by step between mouthfuls.
Browse	Eating of shrub and tree foliage
Drink	Lowers head put mouth to water and swallows.
Submissive	Individual readily retreats if threatened, tail withdrawn.
Crib bite/Wind suck	Horse grabs a fixed object with incisors and pulls back, drawing air into the cranial oesophagus while emitting a characteristic grunt (McGreevy & Nicol, 1998). Also known as cribbing and wind sucking.
Flehmen	Lip curling - a behaviour often associated with stimulation by exciting smells such as sexual hormones
Vocal	Horse makes an audible sound, e.g. whinnie, nicker, and neigh
Urinate	Individual stands with rear legs spread apart and empties bladder
Defecate	Individual lifts tail and empties bowels

Appendix 4: Personality component scores (calculated during principal component analysis of Horse Personality Questionnaire data) of the 44 horses rated reliably in Chapter 3

Horse	Antagonism	Anxiousness	Activity	Protection	Sociability	Inquisitiveness
1	0.14	0.70	-1.30	-0.75	0.97	-1.01
2	-1.33	-0.08	0.84	-1.52	1.19	0.81
3	0.49	0.07	1.20	0.31	0.48	1.07
4	-0.52	-1.24	2.31	0.39	1.86	1.00
5	-0.77	-1.52	0.86	1.09	0.49	0.01
6	-0.89	-0.58	-1.16	1.27	-1.32	-0.09
8	2.18	1.01	0.97	0.47	1.97	-0.53
9	-1.58	-0.26	-1.32	-0.24	1.76	1.59
10	0.27	1.28	0.37	0.30	-0.62	0.87
11	-1.44	2.50	1.91	0.61	-1.72	-0.12
12	-1.71	1.68	0.92	-0.01	0.21	-0.65
13	-0.76	1.87	0.95	0.07	0.59	-1.88
16	-0.39	-0.77	-0.21	0.44	1.57	0.92
17	-2.02	-0.43	1.96	0.48	-0.77	0.00
20	0.45	-0.97	-0.92	-2.19	-0.13	0.52
21	0.62	-1.56	0.82	-0.24	0.35	-0.67
23	1.34	-1.20	1.06	-1.02	-1.06	-1.11
24	-0.01	-0.94	-0.05	0.39	-0.58	0.13
26	-0.58	-0.90	-0.36	-0.12	0.15	1.27
29	-0.38	-1.07	0.22	-0.02	-0.67	-1.11
30	1.42	-0.67	0.07	0.10	-1.00	-0.76
31	0.11	0.63	-0.41	0.49	-1.25	0.13

Appendix 4 (continued): Personality component scores (calculated during principal component analysis of Horse Personality Questionnaire data) of the 44 horses rated reliably in Chapter 3

Horse	Antagonism	Anxiousness	Activity	Protection	Sociability	Inquisitiveness
32	-0.07	-0.44	0.29	0.75	-0.97	0.30
33	0.67	-0.23	0.83	-0.27	-0.71	-0.07
34	0.43	-0.10	0.31	1.22	-1.14	0.40
35	-0.19	0.67	-1.64	1.08	0.35	0.17
37	-0.17	-1.19	0.54	0.48	-1.19	-0.03
38	-0.82	-0.40	-0.30	-0.67	0.18	-0.67
39	0.08	-1.09	-0.61	1.19	1.88	-3.98
41	-0.73	-0.21	-0.77	0.78	-0.38	0.38
42	-0.93	-0.81	-0.20	-1.96	0.76	0.77
43	-0.65	1.55	-0.10	-2.73	0.46	-1.48
48	2.59	1.52	0.47	1.33	1.67	1.57
49	0.56	0.02	-1.59	2.15	0.21	-0.02
51	0.02	0.72	-0.11	0.19	0.50	0.03
52	-0.07	-0.44	0.29	0.75	-0.97	0.30
53	0.67	-0.23	0.83	-0.27	-0.71	-0.07
54	0.43	-0.10	0.31	1.22	-1.14	0.40
55	-0.19	0.67	-1.64	1.08	0.35	0.17
56	-0.17	-1.19	0.54	0.48	-1.19	-0.03
57	-0.82	-0.40	-0.30	-0.67	0.18	-0.67
58	0.08	-1.09	-0.61	1.19	1.88	-3.98
60	-0.73	-0.21	-0.77	0.78	-0.38	0.38
61	-0.93	-0.81	-0.20	-1.96	0.76	0.77

Appendix 5a: Original principal components analysis (varimax rotation) on Rater 2 data (44 horses)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.182	24.729	24.729	6.182	24.729	24.729	4.110	16.439	16.439
2	4.686	18.744	43.473	4.686	18.744	43.473	3.948	15.791	32.231
3	2.428	9.711	53.184	2.428	9.711	53.184	3.655	14.619	46.850
4	1.941	7.763	60.947	1.941	7.763	60.947	2.931	11.723	58.573
5	1.480	5.919	66.866	1.480	5.919	66.866	1.515	6.061	64.634
6	1.380	5.521	72.387	1.380	5.521	72.387	1.464	5.856	70.490
7	1.214	4.855	77.242	1.214	4.855	77.242	1.414	5.656	76.147
8	1.028	4.110	81.352	1.028	4.110	81.352	1.301	5.205	81.352
9	.764	3.055	84.407						
10	.700	2.801	87.209						
11	.613	2.453	89.661						
12	.508	2.031	91.693						
13	.451	1.803	93.495						
14	.354	1.418	94.913						
15	.253	1.011	95.924						
16	.217	.867	96.791						
17	.169	.676	97.467						
18	.159	.634	98.101						
19	.133	.533	98.634						
20	.097	.389	99.023						
21	.082	.326	99.349						
22	.066	.264	99.613						
23	.041	.165	99.779						
24	.041	.164	99.943						
25	.014	.057	100.000						

Appendix 5b: Rotated component matrix for the original principal components analysis (varimax rotation) on Rater 2 data (44 horses)

Adjective	Component							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Active	.059	.192	.102	.883	.156	.155	.065	.158
Aggressive	.704	-.124	-.414	-.009	.194	.055	-.344	.106
Apprehensive	.046	.839	.115	.174	-.071	.225	.146	-.030
Curious	.009	-.086	-.151	.082	.876	-.100	.191	-.002
Eccentric	.786	.216	-.037	.032	-.112	.081	.148	.142
Effective	.186	.020	.210	.209	-.086	.781	.029	.119
Equable	-.518	-.286	.116	-.361	.267	.135	-.146	.357
Excitable	.335	.512	.085	.653	-.012	-.210	-.001	.092
Fearful	.165	.850	.145	.059	.152	-.152	.177	.054
Insecure	-.057	.717	.058	.216	-.035	-.143	.047	-.296
Intelligent	.034	.098	.153	.440	-.033	.120	.203	.764
Irritable	.853	.195	.168	.228	.067	-.071	.012	-.085
Motherly	-.141	-.053	.758	-.039	-.071	-.471	.113	.179
Opportunistic	.204	.087	.014	.049	.174	.005	.806	.123
Playful	-.038	.146	.445	.494	.514	.182	-.278	-.255
Popular	-.091	-.030	.874	.180	.036	.201	-.083	-.072
Protective	.152	.023	.839	-.057	-.074	.174	.095	-.015
Reliable	-.727	.203	.291	-.257	.000	-.082	-.222	.032
Slow	-.145	-.212	.075	-.848	.011	-.126	-.029	-.066
Sociable	-.274	.334	.615	.010	.323	.287	-.058	.081
Stubborn	.741	.342	-.006	-.171	.042	.330	.181	-.065
Subordinate	-.502	.303	.191	-.197	.276	-.164	-.403	.191
Suspicious	.016	.733	-.221	.147	-.048	.166	-.202	.219
Tense	.437	.729	-.026	.242	-.104	.065	-.225	.161
Understanding	-.322	.022	.707	.008	-.240	-.189	-.141	.411

Appendix 6a: Output for principal components analysis (Varimax rotation) on Rater 2 data (44 horses) extraction fixed at six components.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.182	24.729	24.729	6.182	24.729	24.729	4.413	17.653	17.653
2	4.686	18.744	43.473	4.686	18.744	43.473	4.122	16.490	34.143
3	2.428	9.711	53.184	2.428	9.711	53.184	3.663	14.652	48.794
4	1.941	7.763	60.947	1.941	7.763	60.947	2.830	11.321	60.115
5	1.480	5.919	66.866	1.480	5.919	66.866	1.637	6.550	66.665
6	1.380	5.521	72.387	1.380	5.521	72.387	1.430	5.721	72.387
7	1.214	4.855	77.242						
8	1.028	4.110	81.352						
9	.764	3.055	84.407						
10	.700	2.801	87.209						
11	.613	2.453	89.661						
12	.508	2.031	91.693						
13	.451	1.803	93.495						
14	.354	1.418	94.913						
15	.253	1.011	95.924						
16	.217	.867	96.791						
17	.169	.676	97.467						
18	.159	.634	98.101						
19	.133	.533	98.634						
20	.097	.389	99.023						
21	.082	.326	99.349						
22	.066	.264	99.613						
23	.041	.165	99.779						
24	.041	.164	99.943						
25	.014	.057	100.000						

Appendix 6b: Rotated component matrix for the principal components analysis (varimax rotation) on Rater 2 data (44 horses) and restricted to six components.

Adjective	Component					
	1	2	3	4	5	6
Active	.095	.224	.098	.868	.271	.078
Aggressive	.564	-.066	-.421	-.026	.111	.320
Apprehensive	.107	.816	.161	.158	-.019	.078
Curious	.004	-.087	-.153	.027	.824	-.205
Eccentric	.778	.233	-.010	.079	-.150	.011
Effective	.300	-.023	.320	.289	-.032	.576
Equable	-.568	-.353	.163	-.193	.070	.140
Excitable	.279	.578	.033	.599	.057	-.142
Fearful	.140	.847	.147	.045	.113	-.235
Insecure	-.042	.743	.025	.068	.090	-.135
Intelligent	.035	.058	.207	.717	-.185	-.059
Irritable	.796	.266	.136	.140	.116	-.026
Motherly	-.210	-.043	.695	.027	-.141	-.497
Opportunistic	.392	.026	.072	.137	.166	-.505
Playful	-.084	.201	.413	.312	.653	.285
Popular	-.085	-.023	.872	.151	.114	.144
Protective	.179	.012	.858	-.033	-.050	.023
Reliable	-.782	.167	.284	-.227	-.067	.015
Slow	-.184	-.256	.088	-.803	-.134	-.102
Sociable	-.279	.297	.658	.037	.289	.190
Stubborn	.787	.334	.059	-.179	.036	.175
Subordinate	-.666	.295	.169	-.160	.122	.077
Suspicious	-.045	.726	-.187	.196	-.112	.269
Tense	.335	.766	-.023	.245	-.135	.219
Understanding	-.420	.010	.684	.170	-.362	-.127

Appendix 7a: Output for principal components analysis (varimax rotation) on Rater 3 data (44 horses)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.428	25.710	25.710	6.428	25.710	25.710	4.696	18.784	18.784
2	4.383	17.532	43.242	4.383	17.532	43.242	3.492	13.967	32.751
3	3.164	12.655	55.897	3.164	12.655	55.897	3.102	12.409	45.160
4	1.546	6.182	62.079	1.546	6.182	62.079	2.347	9.386	54.546
5	1.418	5.673	67.751	1.418	5.673	67.751	2.264	9.054	63.600
6	1.026	4.105	71.856	1.026	4.105	71.856	2.064	8.256	71.856
7	.960	3.841	75.697						
8	.907	3.629	79.326						
9	.760	3.041	82.367						
10	.611	2.444	84.811						
11	.584	2.335	87.145						
12	.470	1.879	89.024						
13	.416	1.663	90.687						
14	.407	1.628	92.315						
15	.355	1.420	93.735						
16	.316	1.262	94.997						
17	.273	1.091	96.088						
18	.265	1.061	97.149						
19	.178	.713	97.863						
20	.160	.640	98.503						
21	.129	.515	99.018						
22	.096	.382	99.401						
23	.071	.285	99.685						
24	.048	.190	99.876						
25	.031	.124	100.000						

Appendix 7b: Rotated component matrix for principal components analysis on Rater 3 data (44 horses)

Adjective	Component					
	1	2	3	4	5	6
Active	-.116	.321	.341	.087	.500	.537
Aggressive	-.795	.061	.088	-.185	-.029	.184
Apprehensive	.093	.829	.193	-.260	.024	.107
Curious	-.038	-.300	.151	.024	.765	.098
Eccentric	-.276	.000	.610	.054	.404	.148
Effective	-.429	-.181	.315	.161	.225	.546
Equable	.361	-.410	-.583	.054	.034	.218
Excitable	-.094	.414	.546	-.362	.231	.350
Fearful	.033	.899	.043	-.043	.036	.021
Insecure	-.158	.874	.086	-.013	.018	-.033
Intelligent	.176	.186	.600	.118	-.069	.515
Irritable	-.763	.143	.362	-.087	-.017	.080
Motherly	.588	.015	.307	.416	.074	.178
Opportunistic	-.371	-.020	-.073	-.128	.211	.742
Playful	.076	.136	.291	.299	.584	-.024
Popular	.264	-.113	-.101	.671	.328	.040
Protective	.273	-.100	.328	.773	.082	.115
Reliable	.651	-.316	-.311	.287	-.316	.050
Slow	-.085	-.052	-.750	-.069	-.159	-.042
Sociable	.162	.374	-.163	.217	.698	.236
Stubborn	-.748	.041	.052	-.115	-.083	.385
Subordinate	.860	.073	.140	-.223	.050	-.063
Suspicious	.000	.270	.204	-.619	-.142	.382
Tense	-.146	.492	.504	-.087	-.060	.049
Understanding	.768	.056	.050	.369	-.008	.007

Appendix 8: Version of the Horse Personality Questionnaire used in Chapters 4 and 5.

Horse Personality Questionnaire

By Adele Lloyd

In association with:

Moulton College

Open University

Harper Adams University College

Introduction

This questionnaire is an important part of a three-year doctoral research project. Having previously tested this questionnaire on horses and found that it can successfully measure horse personality, this next stage of research explores the existence of personality differences between breeds of horses. In order to complete this research, I am asking owners of various types and breeds of horses to kindly fill in and return this questionnaire.

Instructions

Please read the following guide on how to complete the questionnaire.

- The questionnaire should be filled in by someone who has regularly handled the horse (i.e. Minimum of twice a week) for a minimum of 6 months

- Answers should reflect your opinion of the horse's personality.

- Try to be as accurate as possible when describing the horse's personality. Give the horse the 'score' you most think best describes it.

- If you have more than one pure-bred horse it would be greatly appreciated if you could copy the questionnaire (or contact Adele Lloyd to request more) and fill in for your additional horses.

- Also if you own a pure bred horse of one of the breeds below, it would be greatly appreciated if you could spend some time filling in additional questionnaires for these animals. Stating what breed they are. Thank you.

Irish draught horse, Welsh section A,B,C or D, thoroughbred, Highland pony, Shetland pony, Appaloosa, Arab, American Quarter Horse

Thank you for your cooperation and time, Adele Lloyd

ABOUT YOU:

- Personal questions marked with a * are of particular importance to the study,
- Personal details will only be used in this study and will not be passed onto any other organisations. Your data will be used anonymously as part of an aggregated data base.

Name:.....

*Age: ..... *Gender: .....

*How long have you known the horse? .....

Are you the current owner of the horse? If not please state your connection:

.....

ABOUT THE HORSE:

Name: .....

*Age: ..... *Sex:.....

*Breed .....

*Colour: .....

What is the horse used for? E.g. showing, leisure riding etc

.....

.....

.....

- Please answer the following questions, starting on the next page, using a scale of 1 to 7 (please see the example below)
- Where 1 infers no expression and 7 infers extreme/total expression.
- Choose the number that you think best describes the horse's personality
- Please read each question thoroughly and give some thought to your answer

EXAMPLE

Is the horse ACTIVE? (Moves around a lot and doesn't like being still for long).

Low 1 2 3 4 5 6 7 High

1 = Not active
2 = very rarely active
3 = scarcely active
4 = 50% time active 50% time inactive.
5 = quite active
6 = regularly active
7 = extremely active

If you think the horse is more active than other horses but not extremely active then circle 6 for regularly active.

Low 1 2 3 4 5 (6) 7 High

Please answer the following questions referring to the horse's personality by circling the most appropriate number

1. Is the horse **ACTIVE**? (Moves around a lot and will not stay still for very long).
   Low 1 2 3 4 5 6 7 **High**

2. Is the horse m **AGGRESSIVE**? (Causes or threatens to cause potential harm to other individuals, i.e. humans, horses or other animals).
   Low 1 2 3 4 5 6 7 **High**

3. Is the horse **APPREHENSIVE**? (Appears to be anxious about everything and fears or avoids any kind of risks).
   Low 1 2 3 4 5 6 7 **High**

4. Is the horse **CURIOUS**? (readily explores new situations or objects).
   Low 1 2 3 4 5 6 7 **High**

5. Is the horse **ECCENTRIC**? (Shows stereotypies, unusual mannerisms and exaggerated behaviour).
   Low 1 2 3 4 5 6 7 **High**

6. Is the horse **EFFECTIVE**? (Dominant individual, will regularly get its own way and has an ability to control the behaviour of others).
   Low 1 2 3 4 5 6 7 **High**

7. Is the horse **EQUABLE**? (Very composed and reacts to others, horse and/or human, in an even and calm manner, and is not easily bothered or worried).
   Low 1 2 3 4 5 6 7 **High**

8. Is the horse **EXCITABLE**? (Over reacts to changes, is highly strung and is easily excited or wound up).
   Low 1 2 3 4 5 6 7 **High**

9. Is the horse **FEARFUL**? (Retreats readily from others or outside disturbances and spooks easily).
   Low 1 2 3 4 5 6 7 **High**

10. Is the horse **INSECURE**? (Hesitates to act alone and might seek reassurance from others and tends to be more confident or settled when with other horses).
    Low 1 2 3 4 5 6 7 **High**

11. Is the horse **INTELLIGIENT**? (Learns new things, [i.e. skills, commands etc.] easily and quickly and may have an ability to problem solve. May also appear to benefit from mental stimulation).
    Low 1 2 3 4 5 6 7 **High**

12. Is the horse **IRRITABLE**? (Reacts negatively with little provocation, is highly eruptive and volatile).

Low 1 2 3 4 5 6 7 **High**

13. Is the horse **MOTHERLY**? (Provides a warm, receptive and secure base for others, is tender and caring).

Low 1 2 3 4 5 6 7 **High**

14. Is the horse **OPPORTUNISTIC**? (Seizes a chance as soon as it arises. Will take advantage of a situation).

Low 1 2 3 4 5 6 7 **High**

15. Is the horse **PLAYFUL**? (Initiates play and joins in when play is solicited).

Low 1 2 3 4 5 6 7 **High**

16. Is the horse **POPULAR**? (Others seek it out as a companion).

Low 1 2 3 4 5 6 7 **High**

17. Is the horse **PROTECTIVE**? (Will prevent harm or possible harm to others, defensive of others, human and/or horse).

Low 1 2 3 4 5 6 7 **High**

18. Is the horse **RELIABLE**? (Can be trusted to do things or behaves well, might also be considered a safe horse to be around).

Low 1 2 3 4 5 6 7 **High**

19. Is the horse **SLOW**? (Moves and rests in a relaxed manner, moves slowly but deliberately and is not easily hurried).

Low 1 2 3 4 5 6 7 **High**

20. Is the horse **SOCIABLE**? (Seeks the companionship of others and rarely separates from the group by choice).

Low 1 2 3 4 5 6 7 **High**

21. Is the horse **STUBBORN**? (Does not give in readily or easily, and is not very co-operative with others, human and/or horse).

Low 1 2 3 4 5 6 7 **High**

22. Is the horse **SUBORDINATE**? (Gives in readily to others and submits easily. Will not put up a fight, gets out of the way quickly).

Low 1 2 3 4 5 6 7 **High**

23. Is the horse **SUSPICIOUS** (of others)? (Doesn't trust others readily, human and/or horse, trusts few/select individuals).

Low 1 2 3 4 5 6 7 **High**

24. Is the horse **TENSE**? (Restrained in posture and movement; carries the body stiffly which, suggests a shrinking tendency as if pulling back to be less conspicuous).

Low 1 2 3 4 5 6 7 **High**

25. Is the horse **UNDERSTANDING**? (Responds in a discriminating and appropriate manner to the behaviour of others. Shows a sense of understanding/comprehension/consideration).

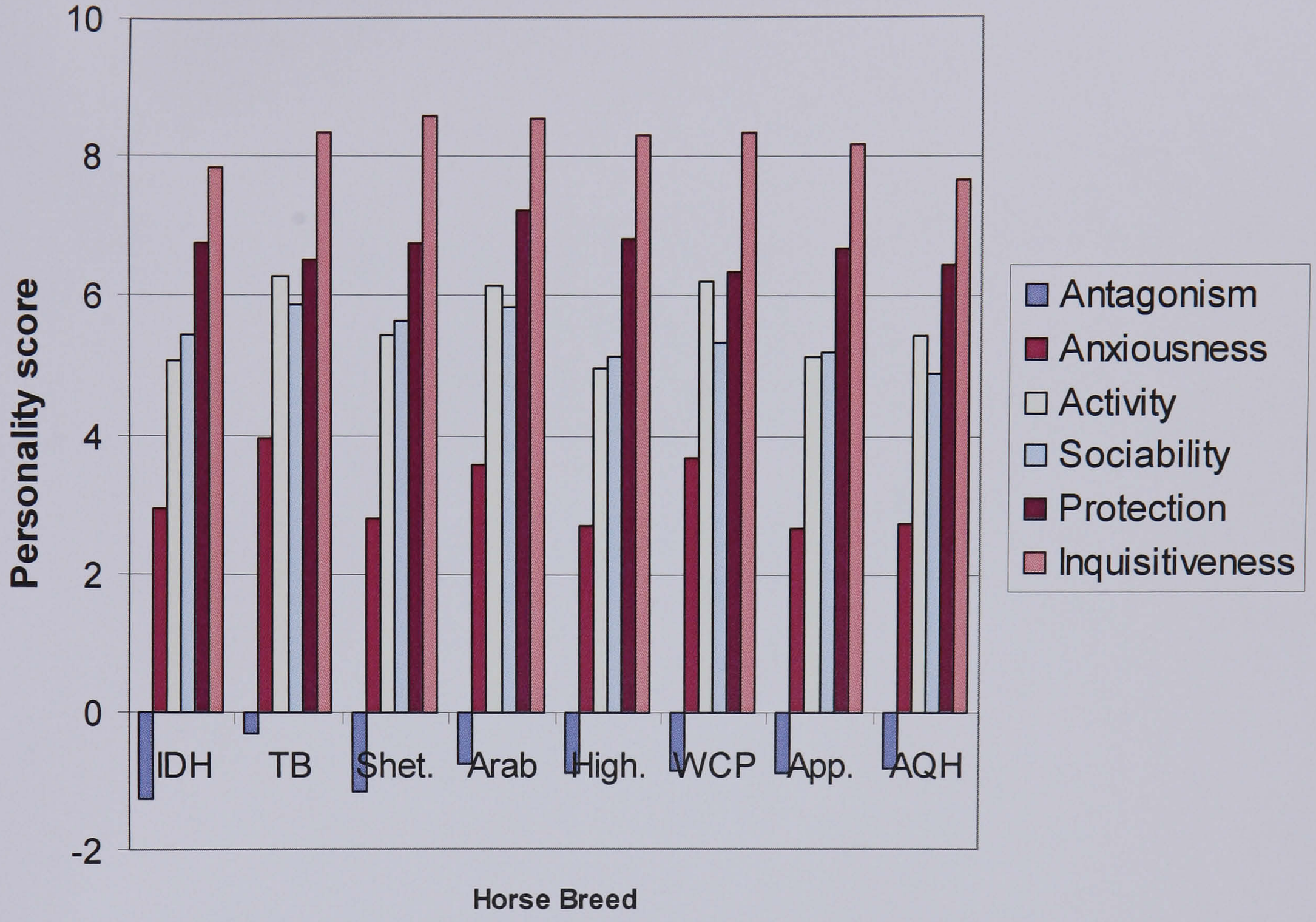
Low 1 2 3 4 5 6 7 **High**

Thank you for completing this questionnaire, if you have any problems or would like more questionnaires or information, please do not hesitate to contact:

Adele Lloyd: Tel. 01604 491131 ext. 608 or email
adeleL@moulton.ac.uk
Moulton College, Moulton, Northampton, NN3 7RR

Once completed please return the questionnaire using the enclosed pre-paid envelope, or post to the above address for the attention of Adele Lloyd.

Appendix 9. Comparison of personality component scores across the eight breeds selected in Chapter 4



IDH = Irish draught horse; TB = thoroughbred; Shet. = Shetland pony; High. = Highland pony; WPC = Welsh ponies and cobs; App. = Appaloosa; AQH = American quarter horse